The 22nd International Conference on Computers in Education

Workshop Proceedings

NOV 30 - DEC 4, 2014 NARA, JAPAN
Editors

Ying-Tien WU
Thepchai SUPNITHI
Tomoko KOJIRI
Chen-Chung LIU
Hiroaki OGATA
Siu Cheung KONG
Akihiro KASHIHARA
Welcome to the Workshop Proceedings of the 22nd International Conference on Computers in Education (ICCE 2014), held from November 29th to December 4th in Nara, Japan.

Established in 1989, ICCE is now an annual international conference organized by the Asia-Pacific Society for Computers in Education (APSCE), and it has become a major event for scholars and researchers in the Asia-Pacific region to share ideas and to discuss their works in the use of technologies in education.

This year we accepted 19 proposals—16 workshops, two interactive events, one tutorial and we aim to explore focused issues in various themes related to the use of technologies in education. Each proposal was peer-reviewed by international reviewers with relevant expertise to ensure high-quality work. All the workshops, organized by the International Program Committees, are in the min-conference format and this proceeding includes all the papers submitted to our workshop. We believe that the workshop provides a valuable opportunity for researchers to share their works and to seek further collaboration. Papers of our workshop will certainly stimulate more interesting research works in these relative areas in Asia-Pacific countries and beyond. We hope that readers will find the newly ideas relevant to their research works in this proceeding.

Finally, we would like to thank the Executive Committee of the Asia-Pacific Society of Computers in Education and the ICCE 2014 International Program Committee Coordination Chair and Co-Chair for entrusting us with this important task of chairing the workshop program and giving us an opportunity to work with many outstanding researchers. We would also like to thank the Local Organizing Committee for helping with the logistic of the workshop program.

Workshop, Tutorial and Interactive Event Coordination Co-Chairs:

*Ying-Tien Wu*, National Central University, Taiwan

*Thepchai Supnithi*, National Electronics and Computer Technology Center, LST, Thailand

*Tomoko Kojiri*, Kansai University, Japan
# TABLE OF CONTENTS

**Workshop 1: Natural Language Processing Techniques for Educational Applications (NLP-TEA), with a Shared Task on Grammatical Error Diagnosis for Learning Chinese as a Foreign Language (CFL)**

1. A Sentence-Pattern Learning Support System for Japanese  
   *Takahiro Ohno, Ayato Inoue & Dongli Han*  
   1

2. CYCCDC: A Chiayi Chinese Conversation Dialogue Corpus  
   *Jui-Feng Yeh, Yun-Yun Lu & Yi-Syun Tan*  
   7

3. Partial and Synchronized Caption Generation to Develop Second Language Listening Skill  
   *Maryam Sadat MIRZAEI, Yuya AKITA & Tatsuya KAWAHARA*  
   13

4. Challenges in the Annotation of Article Errors in Spanish Learner Texts  
   *María del Pilar VALVERDE IBÁÑEZ*  
   24

5. Tools for Supporting Language Acquisition via Extensive Reading  
   *Alexandra UITDENBOGERD*  
   35

6. Overview of Grammatical Error Diagnosis for Learning Chinese as a Foreign Language  
   *Liang-Chih YU, Lung-Hao LEE & Li-Ping CHANG*  
   42

7. KNGED: a Tool for Grammatical Error Diagnosis of Chinese Sentences  
   *Tao-Hsing CHANG, Yao-Ting SUNG, Jia-Fei HONG & Jen-I CHANG*  
   48

8. Extracting a Chinese Learner Corpus from the Web: Grammatical Error Correction for Learning Chinese as a Foreign Language with Statistical Machine Translation  
   *Yinchen ZHAO, Mamoru KOMACHI & Hiroshi ISHIKAWA*  
   56

9. Detecting Grammatical Error in Chinese Sentence for Foreign  
   *Jui-Feng Yeh, Yun-Yun Lu, Chen-Hsien Lee, Yu-Hsiang Yu & Yong-Ting Chen*  
   62

10. Grammatical Error Detection with Limited Training Data: The Case of Chinese  
    *Marcos ZAMPIERI & Liling TAN*  
    69
11. Description of NTOU Chinese Grammar Checker in CFL 2014
   Chuan-Jie Lin & Shao-Heng Chan

**Workshop 2: Third International Workshop on ICT Trends in Emerging Economies**

12. Harnessing ICT for Educational Development in Emerging Developing Countries within the Asia-Pacific Region
   Mas Nida MD KHAMBARI, Ahmad Fauzi MOHD AYUB & Mohammad LUKMAN

13. Learning and Affect Trajectories Within Newton’s Playground
   Juan Miguel L. ANDRES & Ma. Mercedes T. RODRIGO

   Sousan BALEGHI ZADEH, Ahmad Fauzi MOHD AYUB, Rosnaini MAHMUD & Shaffee MOHD DAUD

15. Exploring Deep Approach to Learning for Accounting through ICT-Supported Learning Environment in Malaysian Secondary Schools: A Preliminary Study
   Boon See TAN & Su Luan WONG

16. Technology-enhanced Chemistry Learning and Students' Perceptions: A Comparison of Microcomputer-based Laboratory and Web-based Inquiry Science Environment
   Kulthida KAMTOOM & Niwat SRISAWASDI

17. The Design of Instructional Scaffolds to Facilitate Online Project-Based Learning
   Chun-Ping WU, Ching-Chiu YEH, Shu-Ling WU & Hao Jie YONG

18. GeMA-ICT Learning Effectiveness in Improving Student Mathematical Ability
   Sigid Edy PURWANTO, Wahidin & Aidiyah Novian NISYAH

19. The Effectiveness of Association Picture Media Applications in Katakana Letter Reading Comprehension of Grade Ten Students of SMK Manajemen (Vocational Management School, Jakarta)
   Restoe NINGROEM, Endy SJAIFUL ALIM
20. An Online Survey: Studying the Antecedents of Technology Use Through the UTAUT Model Among Arts and Science Undergraduate Students
   Priscilla MOSES, Tiny Chiu Yuen TEY, Phaik Kin CHEAH, Timothy TEO & Su Luan WONG

Workshop 3: Computer-Supported Visualization, Modeling, and Simulations for Learning

21. Equipping High School Students with the Abilities of Evaluating Evidence and Formulating Evidence for an On-line Decision-making Task
   Shu-Sheng LIN & Ying-Shao HSU

22. An Investigation of Relationships between Biology Attitudes and Perceptions toward Instructional Technology in Analogy-based Simulation on Light Reaction
   Sarunya PINATUWONG & Niwat SRISAWASDI

   Wei-Yan LIN, Meng-Jung TSAI & An-Hsuan WU

   An-Hsuan WU, Po-Fen HSU, Hui-Jou CHIU & Meng-Jung TSAI

25. Incorporating Augmented Reality into Learning Practical Skills for Medical Surgery
   Ying-Shao Hsu, Yuan-Hsiang Lin, Beender Yang, Shih-Fan Yang, Ya-Yen Chan, Zi-Hao Lin & Yi-Sheng Chan

26. Path Analyses of How Students Develop Conceptual Knowledge and Inquiry Skills in a Simulation-Based Inquiry Environment
   Hsin-Yi CHANG, Ying-Shao HSU, Hsin-Kai WU & Chih-Ming CHEN

27. Understanding Middle and High School Students' views of Model Evaluation and Model Change
   Silvia Wen-Yu Lee, Hsin-Kai WU & Hsin-Yi Chang

Workshop 4: The 3rd Workshop on Application of Innovative Educational Technologies in STEM Education

28. Bio Detective: Student Science Learning, Immersion Experience, and Problem-solving Patterns
   Mei-En HSU & Meng-Tzu CHENG
29. The Effect of Students’ Effectiveness and Attitude in Heterogeneous and Free Grouping Cooperative Learning Applied in Sixth-grade Students’ Scratch Program Teaching
Lei CHEN, Xiuyu YANG, Xintong WANG & Feng-kuang CHIANG

30. Designing Mobile Application for STEM: Building Individual Interest and Supporting Creative and Innovative Thinking Skills
Ilker YENGIN

31. The Development and Evaluation of an Educational Game- Shimmer© with Computer Visualization for Optics Learning
Huei-Tse HOU, Shu-Ming WANG, Hsin-Hung YU & Shie-He LIU

32. A Three-Stage Augmented-Reality-Facilitated Earth Science Instructional Process for Dispersing Learning Style Differences
Chang-Hwa WANG & Cheng-Ping CHEN

33. The Effects of AR-based Instruction on Students’ Learning Performance, Motivation and Self-efficacy in Programming Learning
Gloria Yi-Ming KAO & Cheng-An RUAN

34. Implementation of Student-associated Game-based Open Inquiry in Chemistry Education: Results on Students' Perception and Motivation
Keeratika MEESUK & Niwat SRISAWASDI

35. The Development and Evaluation of the Online Science Fair Inquiry System based on Scaffolding Design
Li-jen WANG, Chien-yu, CHENG, Chiu-ming, HU & Ying-Tien WU

36. The Difference in Sudoku Puzzle-Solving Ability Between Undergraduates and Postgraduates
Hong-Mei HU, Ling-Jin LI, Li-Sha WANG & Feng-Kuang CHIANG

37. Investigating the Role of Self-explanation and Co-explanation in 4th Graders’ Game-based Science Learning
Chung-Yuan HSU, Hung-Yuan WANG & Shih-Hsuan WEI

Workshop 5: 1st ICCE Workshop on Learning Analytics (LA): Leveraging Educational Data for Adaptive Learning and Teaching

38. Preliminary Requirements Analysis towards an Integrated Learning Analytics System
Byung-gi CHOI, Yong-sang CHO & Jaeho LEE
<table>
<thead>
<tr>
<th>No.</th>
<th>Title</th>
<th>Pages</th>
</tr>
</thead>
<tbody>
<tr>
<td>39.</td>
<td>Learning Analytics Interoperability – Looking for Low-Hanging Fruits</td>
<td>253</td>
</tr>
<tr>
<td></td>
<td>Tore Hoel &amp; Weiqin Chen</td>
<td></td>
</tr>
<tr>
<td>40.</td>
<td>Making Sense of Online Learning Behavior: A Research on Learning Styles and Collaborative Learning Data</td>
<td>265</td>
</tr>
<tr>
<td></td>
<td>Meng SUN, Jiu-Tong LUO, Dong-Ming QIAN &amp; Xiao-Qing GU</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Christian M. STRACKE</td>
<td></td>
</tr>
<tr>
<td>42.</td>
<td>Learning Analytics Data Items on Digital Textbooks</td>
<td>277</td>
</tr>
<tr>
<td></td>
<td>Yasuhisa Tamura</td>
<td></td>
</tr>
<tr>
<td>43.</td>
<td>Learning Analytics: An Enabler for Dropout Prediction</td>
<td>286</td>
</tr>
<tr>
<td></td>
<td>Shu-Fen TSENG, Chih-Yueh CHOU, Zhi-Hong CHEN &amp; Po-Yao CHAO</td>
<td></td>
</tr>
</tbody>
</table>

**Workshop 6: Emerging Pedagogies for Computer-based Learning**

<table>
<thead>
<tr>
<th>No.</th>
<th>Title</th>
<th>Pages</th>
</tr>
</thead>
<tbody>
<tr>
<td>44.</td>
<td>Motivation and Engagement in MOOC – Teachers’ Perspective</td>
<td>289</td>
</tr>
<tr>
<td></td>
<td>Li Fern TAN, Kai Song GOH &amp; Emile SABASTIAN</td>
<td></td>
</tr>
<tr>
<td>45.</td>
<td>Effects of Gender Differences and Learning Performance within Residence Energy Saving Game-based Inquiry Playing</td>
<td>298</td>
</tr>
<tr>
<td></td>
<td>Ugyen DORJI, Patcharin PANJABUREE &amp; Niwat SRISAWASDI</td>
<td></td>
</tr>
<tr>
<td>46.</td>
<td>A Blended Learning Environment in Chemistry for Promoting Conceptual Comprehension: A Journey to Target Students' Misconceptions</td>
<td>307</td>
</tr>
<tr>
<td></td>
<td>Sumarin NIROJ &amp; Niwat SRISAWASDI</td>
<td></td>
</tr>
<tr>
<td>47.</td>
<td>Investigating Correlation between Attitude toward Chemistry and Motivation within Educational Digital Game-based Learning</td>
<td>316</td>
</tr>
<tr>
<td></td>
<td>Nattida NANTAKAEW &amp; Niwat SRISAWASDI</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Chiu-Ming Hu, Chao-Shen Cheng, Li-Jen Wang, Huei-Tse Hou, Yi-Chun Kuo, Cheng Teng Yao &amp; Ying-Tien Wu</td>
<td></td>
</tr>
<tr>
<td>No.</td>
<td>Title</td>
<td>Authors</td>
</tr>
<tr>
<td>-----</td>
<td>----------------------------------------------------------------------</td>
<td>-------------------------------------------------------------------------</td>
</tr>
<tr>
<td>49</td>
<td>An Evaluation of Macro-Micro Representation-based Computer Simulation</td>
<td>Jarunya BUYAI &amp; Niwat SRISAWASDI</td>
</tr>
<tr>
<td></td>
<td>for Physics Learning in Liquid Pressure: Results on Students’ Perceptions and Attitude</td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>Promoting Students’ Physics Motivation by Blended Combination of</td>
<td>Chakkrapan PIRAKSA &amp; Niwat SRISAWASDI</td>
</tr>
<tr>
<td></td>
<td>Physical and Virtual Laboratory Environment: A Result on Different Levels of Inquiry</td>
<td></td>
</tr>
<tr>
<td>51</td>
<td>An Experimental Study on the Effects of an Online Student-Constructed Tests Learning Activity</td>
<td>Fu-Yun YU &amp; Chia-Ling SU</td>
</tr>
<tr>
<td>52</td>
<td>Exploring the Effects of Student Question-Generation Strategy</td>
<td>Chun-Ping WU, Shu-Ling WU &amp; Ching-Chiu YEH</td>
</tr>
<tr>
<td>53</td>
<td>Structured Explanation Generation for Conceptual Understanding in Physics</td>
<td>Tomoya HORIGUCHI, Takahito TOUMOTO &amp; Tsukasa HIRASHIMA</td>
</tr>
<tr>
<td>54</td>
<td>Practical Use of Interactive Environment for Learning by Problem-posing for One-step Multiplication and Division Word Problems</td>
<td>Sho YAMAMOTO, Yuki AKAO, Mitsutaka MUROTSU, Takehiro KANBE, Yuta YOSHIDA, Kazushige MAEDA, Yusuke HAYASHI &amp; Tsukasa HIRASHIMA</td>
</tr>
<tr>
<td>55</td>
<td>Revealing Students’ Thinking Process in Problem-Posing Exercises:</td>
<td>Nur HASANAH, Yusuke HAYASHI &amp; Tsukasa HIRASHIMA</td>
</tr>
<tr>
<td></td>
<td>Analysis of First Sentence Selection</td>
<td></td>
</tr>
<tr>
<td>56</td>
<td>Balance Control of Question-Posing Focusing on Learning Target Words</td>
<td>Toshihiro HAYASHI, Yuiji HIRAI, Kazuhiro URA, Akihiro IWAKI, Rihito YAEGASHI, Hiroshi MURAI &amp; Hiroyuki TARUMI</td>
</tr>
<tr>
<td></td>
<td>on the Self-Study Material Contribution and Sharing System</td>
<td></td>
</tr>
</tbody>
</table>
Workshop 8: Computer-Supported Personalized and Collaborative Learning

57. Development of a Customized English Learning System based on Augmented Reality Technology
   Gwo-Haur HWANG, Beyin CHEN & Hen-Lin HUANG

58. Development of a Multi-Device Data Structures Course Item Bank Practice System with Self-Regulated Learning Strategy on Bloom's Taxonomy of Educational Objectives
   Gwo-Haur HWANG, Jing-Fang CHEN, Yu-Ting SHIH, Yong-Sheng JHANG, Yi-Xuan LIN & Yu-Syuan WANG

59. Students' Self-efficacy and Acceptance toward Context-Aware Ubiquitous Learning in Biology Education: A Case of Photosynthesis in Plant
   Chuntanet NASARO & Niwat SRISAWASDI

60. How Self-Efficacy Affects Students’ Performance and Pace in Self-Directed Learning with ICT
   Andrew C.-C. LAO, Mark C.-L. HUANG, Oskar KU & Tak-Wai CHAN

61. The Effects of Game-based Peer Response on Writing Quality: High-ability vs. Low-ability
   Jen-Hang WANG, Sherry Y. CHEN, Oskar KU & Tak-Wai CHAN

62. The Effects of Mini-Games on Students’ Confidence and Performance in Mental Calculation
   Oskar KU, Denise H. WU, Andrew C. C. LAO, Jen-Hang WANG & Tak-Wai CHAN

63. The Interface Design of Electronic Journals via Mobile Devices: A Cognitive styles Perspective
   Chu-Han CHAN & Sherry Y. CHEN

64. Enhancing Metacognition through Weblog in Physics Classroom Thai Context
   Jirutthitikan PIMVICHAI, Chokchai YUENYONG & Sakanan ANANTASOOK

65. Knowledge Propagation in Practical Use of Kit-Build Concept Map System in Classroom Group Work for Knowledge Sharing
   Toshihiro NOMURA, Yusuke HAYASHI, Takuma SUZUKI & Tsukasa HIRASHIMA
<table>
<thead>
<tr>
<th></th>
<th>Title</th>
<th>Authors</th>
</tr>
</thead>
<tbody>
<tr>
<td>66</td>
<td>The Exploration of Improving Efficiency of Synchronous Discussion: e-</td>
<td>I-Fan LIU &amp; Chun-Wang WEI</td>
</tr>
<tr>
<td></td>
<td>Case Live Show</td>
<td></td>
</tr>
<tr>
<td>67</td>
<td>Game playing as a Strategy to Improve Team Cohesion, Support for</td>
<td>Pet-Yu CHENG, Wen-Yen WANG &amp; Yueh-Min HUANG</td>
</tr>
<tr>
<td></td>
<td>Collaborative U-Learning</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>**Workshop 9: The Applications of Information and Communication</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Technologies in Adult and Continuing Education</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>68</td>
<td>The Relationship Between Parents Addicted to Mobile Phone and</td>
<td>Ying ZHOU, Xiao ZHANG, Jyh-Chong LIANG &amp; Chin-Chung TSAI</td>
</tr>
<tr>
<td></td>
<td>Adolescent Addicted to the Internet</td>
<td></td>
</tr>
<tr>
<td>69</td>
<td>Developing an Instrument to Assess Teachers’ Belief, Confidence and</td>
<td>Yu-Hsuan CHANG &amp; Meng-Jung TSAI</td>
</tr>
<tr>
<td></td>
<td>Motivation about Digital Game-based Learning</td>
<td></td>
</tr>
<tr>
<td>70</td>
<td>Eye-tracking Analyses of Text-and-graphic Design Effects on E-book</td>
<td>Tse-Wen PAN &amp; Meng-Jung TSAI</td>
</tr>
<tr>
<td></td>
<td>Reading Process and Performance: “Spanish Color Vocabulary ” as an</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Example</td>
<td></td>
</tr>
<tr>
<td>71</td>
<td>Weblog as Learning Community for Supporting Astronomy Teaching in</td>
<td>Sakanan ANANTASOOK &amp; Chokchai YUENYONG</td>
</tr>
<tr>
<td></td>
<td>Thailand</td>
<td></td>
</tr>
<tr>
<td>72</td>
<td>The Perceived and Expected User Experiences of AR Book Reading: the</td>
<td>Kun-Hung CHENG &amp; Chin-Chung TSAI</td>
</tr>
<tr>
<td></td>
<td>Perspective of Parents</td>
<td></td>
</tr>
<tr>
<td>73</td>
<td>Design of MOOC for In-service Teacher Professional Development: A</td>
<td>Silu LI, Eric T. H. LUK &amp; Morris S. Y. JONG</td>
</tr>
<tr>
<td></td>
<td>Case of Teachers’ Refresher Training Course in Hong Kong</td>
<td></td>
</tr>
<tr>
<td>74</td>
<td>Learning to Create Technological Pedagogical Content Knowledge through</td>
<td>Ching Sing CHAI, Benjamin WANG &amp; Chun Ming TAN</td>
</tr>
<tr>
<td></td>
<td>Distributed Leadership: A Case Study of a Singapore Future School</td>
<td></td>
</tr>
<tr>
<td>75</td>
<td>Development and Validation of an Instrument for Exploring Taiwanese</td>
<td>Yu-Chih TSAO, Chi-Ling WU &amp; Min-Hsien LEE</td>
</tr>
<tr>
<td></td>
<td>Undergraduates’ Approaches to Internet-based Learning</td>
<td></td>
</tr>
</tbody>
</table>
76. Investigating Chinese University Students’ Perceptions about Blackboard Platform to Support Their Online Learning
   Weisheng LI, Lizhu CHEN, Ge QU & Yan DONG

77. Engineering Graduate Students’ Literature Searching Behaviors
   Ying-Hsueh CHENG & Chin-Chung TSAI

78. Developing an Online Formative Assessment System for a Chinese EFL Course
   Chunping ZHENG, You SU & Jingjing LIAN

79. Role-Play in Computer-Supported Collaborative Learning-An Explorative Study
   Yu-Chen HSU & Yen-Lin CHIU

80. Exploring the Interactive Use of Video Cases in Scaffolding Prospective Teachers in Learning Clinical Interview Method
   Yu-Ling HSU


81. The Effectiveness of Reducing State Anxiety by Digital Counseling Tool-Mind Collage
   Yu-Jen Hsu & Ju-Ling Shih

82. Math Island: Designing a Management Game of Primary Mathematics for Facilitating Student Learning
   Charles Yen-Cheng Yeh, Hercy N.H. Cheng, Zhi-Hong Chen & Tak-Wai Chan

83. The Change of Interpersonal Relationship for Group Development in Digital Game-based Adventure Education Course
   Chang-Hsin LIN, Yu-Jen Hsu, Ju-Ling Shih & Chia-Chun TSENG

84. Designing Educational Computer Game for Human Circulatory System: a Pilot Study
   Jatuput LOKAYUT & Niwat SRISAWASDI

85. The Curriculum Design of Nutrition and Food Safety Game for Elementary School Student
   Chun-Heng LIN, Tsung-Yen CHUANG, Chung-Chiann CHUANG, Fang-Ying TU & Hua-Hsiang TSENG
86. Thinking as a Pleasure: Tactics to Design Digital Educational Games from the Perspective of Board Games
Hercury Nien-Heng CHENG 585

87. VocaMono: An Online Multiplayer English Vocabulary Learning Board Game
Jia-Jiunn Lo & Chin-Kun Hsin 591

88. The Application of Game-Based Learning in Early Childhood Acquisition
Sanko Lan, Joni Tzuchen Tang & Yie-Su HWANG 597

Workshop 11: International Workshop on "Technology Enhanced Language Learning"

89. Correlation of Professional English Reading VS. Eye Gazing and Frequency of Rereading Eye Movement
Hong-Fa HO, Guan-Yu HOU, Chen-Ku LIN, Chen-Hsiung LIN & Soh O-K 601

90. Assisting Tools for Selecting Proper Semantic Meaning by Disambiguation of the Interference of the First Language
Nattapol KRITSUTHIKUL, Shinobu HASEGAWA, Cholwich NATTEE & Thepchai SUPNITHI 609

91. Effects of Students Using Smartphones to Receive Different Amount of L1 Support for Listening Comprehension and Vocabulary Recall
Gwo-Jen HWANG, Yi-Hsuan HSIEH & Ching-Kun HSU 616

Yu-Ju LAN, Hsiao-Hsuan WEI & Ya-Li CHIU 625

Katrin SAKS & Äli LEIJEN 630

94. The Impacts of Using Interactive E-book on the Learning Effectiveness of English Blank-filling Cloze
Gwo-Jen HWANG, Yi-Hsuan HSIEH, Ching-Jung HSUEH & Ching-Kun HSU 640
95. Correlation of English Test Outcome From TVE Joint College Entrance Examination of Taiwan VS. Professional English Reading Speed and Comprehension
   Hong-Fa HO, Yi-Yeh CHUNG, Chen-Ku LIN, Chen-Hsiung LIN & SOH O-K

96. A Tablet-based Chinese Composition Assessment System
   Kat LEUNG, Barley MAK & Howard LEUNG

97. Learning to Learn Collaboratively on Facebook – A Pilot Study
   Sarah Hsueh-Jui LIU & Yu-Ju LAN

Workshop 12: The 5th International Workshop on "Technology---Transformed Learning: Going Beyond the One-to-One Model?"

98. Exploring the Effectiveness of a Flipped Classroom Based on Control-Value Theory: A Case Study
   Jiu-Tong LUO, Meng SUN, Bian WU & Xiao-Qing GU

99. Analysis on Students’ Acceptance of Digital Reading in Ubiquitous Cooperative Inquiry-based Learning Environment
   Jing-Ya CHEN, Jing LENG, Xiao-Juan XU & Xiao-Qing Gu

100. Visualizing Ubiquitous Learning Logs Using Collocational Networks
    Kousuke Mouri, Hiroaki Ogata, Noriko Uosaki & SongRan Liu

101. The Research of China’s Policies and Practices of Life-long Learning in U-learning Environment
    Bingqian JIANG, Jun XIAO, Jing LENG & Xiaoqing GU

102. Phonic Social Network Software Scaffolds Language Learning in Ubiquitous Learning Environment
    Huawen WANG, Jing LENG & Xiaoqing GU

103. We are going to the ZOO! Virtual Badges in Formal out-of-school 1:1 Learning Journey with Smartphones
    Ivica BOTICKI, Jelena BAKSA, Peter SEOW & Chee-Kit LOOI

104. Building an Online Collaborative Learning Community in Ubiquitous Learning Environment
    Ru ZHANG, Jing LENG, Xiaoqing GU, Guanfeng FU & Huawen WANG
<table>
<thead>
<tr>
<th>No.</th>
<th>Title</th>
<th>Authors</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>105.</td>
<td>Identifying User's Perceptions Toward Integrating Mobile Applications in Science Education</td>
<td>Hyo-Jeong SO, Hye-Gyoung YOON, Hyungshin CHOI, Heung-Chang LEE &amp; Kyudong PARK</td>
<td>726</td>
</tr>
<tr>
<td>106.</td>
<td>Integration of Multiple External Representations in Chemistry: A Requirements-gathering Study</td>
<td>Prajakt PANDE &amp; Sanjay CHANDRASEKHARAN</td>
<td>732</td>
</tr>
<tr>
<td>108.</td>
<td>Using Ontology for Representing Role Change Design in Nursing Service Thinking Education</td>
<td>Wei CHEN, Liang CUI, Kaji TANAKA, Hirotaka NISHIYAMA, Noriyuki MATSUDA &amp; Mitsuru IKEDA</td>
<td>744</td>
</tr>
<tr>
<td>109.</td>
<td>Self-assessment Rubrics as Metacognitive Scaffolds to Improve Design Thinking</td>
<td>Madhuri MAVINKURVE &amp; Sahana MURTHY</td>
<td>750</td>
</tr>
<tr>
<td>110.</td>
<td>The Application of QR Codes in Outdoor Education Activities: Practice and Discussion</td>
<td>Wen-Shian Lee &amp; Chun-Yen Chang</td>
<td>756</td>
</tr>
<tr>
<td>111.</td>
<td>The Application of Instructional Media and IRS in Environmental Education - Focus on the Rocky Terrain in Northern Coast of Taiwan</td>
<td>Wen-Mao Chung, Te-Shin Tsai &amp; Chun-Yen Chang</td>
<td>766</td>
</tr>
<tr>
<td>112.</td>
<td>Effect of Inquiry Web-Based Learning Competition for Gifted Students in Junior High School</td>
<td>Yen-Hung Shen &amp; Wen-Gin Yang</td>
<td>772</td>
</tr>
<tr>
<td>113.</td>
<td>Evaluation of the Situation Somatosensory Game Digital Learning for Global Warming Misconception</td>
<td>Hsin-Chih Lai &amp; Chi-Chen Li</td>
<td>777</td>
</tr>
</tbody>
</table>
114. The Environmental Education of Migration Birds Using a Near Time Web-based Design
Chow-Jeng Wang, Chen-Jeih Pan & Yi Jong Tsai

Workshop 15: The 4th Workshop on Skill Analysis, Learning or Teaching of Skills, Learning Environments or Training

115. Development of Sign Language Training Machine Using Depth Sensor
Yuichiro MORI, Akie FUKUHARA & Shogo HAYASHIDA

116. Significance and Possibility of E-Learning for Choreographic Skills in Contemporary Dance
Bin UMINO, Asako SOGA & Motoko HIRAYAMA

117. Design of an Environment for Motor-skill Development based on Real-time Feedback
Keita YAMADA & Kenji MATSUURA

118. Content Management System to Support Improvement in Quality of Fitness Testing of Athletes
Yuji KOBAYASHI & Naka GOTODA

119. Development of a Learning Environment for Novices' Erhu Playings
Fumitaka KIKUKAWA, Masato SOGA & Hirokazu TAKI

120. Analysis and Feedback of Baseball Pitching Form with Use of Kinect
Yasuhisa TAMURA, Taro MARUYAMA & Takeshi SHIMA

Workshop 17: International Workshop on Innovative Design of Learning Space

121. The Integration of Augmented Reality Mobile Learning and Self-Regulated Learning by Using Concept Mapping - A Case Study of the Plants in Campus
Po-Han WU, Gwo-Haur HWANG, Yu-Syuan WANG & Yen-Ru SHI

122. Application of Teams-Games-Tournament Strategy to Investigate Learning Effectiveness in Primary Schools
Shu-Hsien Huang, Ting-Ting Wu & Yueh-Min Huang

123. Education 3.0 and Beyond: A learner-led Experience of Education
Pranav KOTHARI & Anurima CHATTERJEE
<table>
<thead>
<tr>
<th>124.</th>
<th>The Impact of Affective Tutoring System and Information Literacy on Elementary School Students’ Cognitive Load and Learning Outcomes</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Ching-Ju CHAO, Shang-Chin TSAI, Chia-Hsun LEE, Tao-Hua WANG &amp; Hao-Chiang Koong LIN</td>
</tr>
<tr>
<td>125.</td>
<td>Development of Digital Game-based Biology Learning Experience on Cell Cycle through DSLM Instructional Approach</td>
</tr>
<tr>
<td></td>
<td>Porntip KANYAPASIT &amp; Niwat SRISAWASDI</td>
</tr>
<tr>
<td>126.</td>
<td>Mobile Augmented Reality in Supporting Performance Assessment: An Implementation in a Cooking Course</td>
</tr>
<tr>
<td></td>
<td>Kuo-Hung CHAO, Chung-Hsien LAN, Yao-Tang LEE, Kinshuk, Kuo-En CHANG &amp; Yao-Ting SUNG</td>
</tr>
<tr>
<td>127.</td>
<td>Investigating Effects of Mobile Learning in Familiar Authentic Environment on Learning Achievement and Cognitive Load</td>
</tr>
<tr>
<td></td>
<td>Rustam SHADIEV, Wu-Yuin HWANG, Yueh-Min HUANG &amp; Tzu-Yu Liu</td>
</tr>
<tr>
<td>128.</td>
<td>Applying Adaptive Hybrid Recommendation Technology for Searching Algorithm Learning Articles</td>
</tr>
<tr>
<td></td>
<td>Shu-Chen Cheng &amp; Shih-Che Huang</td>
</tr>
</tbody>
</table>
A Sentence-Pattern Learning Support System for Japanese

Takahiro Ohno*, Ayato Inoue* & Dongli Han**
*Graduate School of Integrated Basic Sciences, Nihon University, Tokyo, Japan
**Department of Information Science College of Humanities and Sciences, Nihon University, Tokyo, Japan

Abstract: Computers have played more and more important roles in Japanese education. We in this paper focus on sentence-patterns in Japanese and have developed a Web system to support sentence-pattern learning. Specifically, our system is able to indicate a sentence pattern from a freely composed Japanese sentence by using a digital sentence-pattern dictionary generated in advance. After that, our system determines whether the sentence pattern has been used properly and gives some instructions on how to use the sentence pattern correctly. Experiments show the effectiveness of our method.

Keywords: Japanese education, learning materials, Web application, sentence pattern, free composition oriented system

1. Introduction

The number of Japanese learners all over the world is increasing every year. According to a report by Japan Foundation (2013), the total number has reached approximate 4 millions spreading in 136 countries, which is 9.2% larger than the total number counted four years before. On the other hand, a parallel investigation towards all the Japanese education institution around the world shows that the lack of learning material and facility, the monotony in learning method, and the learners’ lukewarm attitude still remain as the most serious problems in Japanese learning. Computers and learning support tools based on Web will undoubtedly provide a good solution to the above problems more or less. They might be helpful not only to the learners but also to the teachers as well.

Many e-learning systems have been developed to help users learn Japanese in a simple manner with true-false or fill-in-the-blank questions only (e.g., Samidori1, WebCM2, and Nihongo-Dekimasu3). In contrast, free-composition oriented systems are considered more helpful but more difficult to implement. Here are some existing web systems based on free composition. Obi2 analyzes an input sentence and classifies it into 13 difficulty-levels (Sato etc., 2008). Asunaro conducts dependency-relation analysis on input sentences (Abekawa etc., 2002). Reading Tutor explains each morpheme contained in an input Japanese sentence in six different languages, and discovers the sentence-patterns used (Kawamura, 2002). These systems support free composition and hence make themselves more practical and feasible. However, none of the above systems is capable of handling wrong texts, which is more realistic and hence has more actual significance. Here, a wrong text indicates the text containing grammatically wrong usages.

In this paper, we focus on sentence-patterns in Japanese and aim to develop a free-composition oriented Web system for Japanese learners in all levels. Specifically, our system is able to recognize a sentence pattern from an input Japanese sentence by using a digital sentence-pattern dictionary.

---

1 http://www.ku-japanese.jp/
2 http://opal.ecis.nagoya-u.ac.jp/webcmj/
3 http://www.erin.ne.jp/
generated in advance. After that, our system determines whether the sentence pattern has been used properly and gives some instructions on how to use the sentence pattern correctly. The difference between our system and Reading Tutor is that our system does not expect users to input correct sentences at all, whereas Reading Tutor only accepts sentences containing correctly used sentence patterns.

In the rest of this paper, Section 2 describes the pre-created sentence-pattern dictionary, Section 3 details the discovering and correcting process of sentence-pattern, and Section 4 shows the experimental results for our approach. Finally, we end this paper with a conclusion in Section 5.

2. Sentence-pattern Dictionary

A Japanese sentence pattern is composed of a set of words in a fixed order to express some particular meaning (Han and Song, 2011). A simple example is “〜しだい” meaning as soon as. The symbol “〜” is a placeholder where only expressions satisfying some certain conditions could be inserted. For the sentence pattern “〜しだい” here, only two kinds of expressions could be used to replace “〜” in front of “しだい”:

1. predicative forms of verbs
2. Sahen-verbs which are formed by adding “する” to action nouns

Sentence patterns are supposed to be one of the most difficult issues during the process of learning Japanese. And we consider it is necessary and important to make users aware of their own usage of sentence patterns when composing Japanese documents.

The first step we take to build a learning support system with the capability of discovering and correcting sentence-patterns is generate a sentence-pattern dictionary. Totally, seven types of structures exist in Japanese sentence patterns as shown in Figure 1. As stated above, “〜” in Figure 1 is a placeholder, and each symbol other than “〜” stands for a fixed expression contained in a particular sentence pattern. For example, the sentence pattern “〜しだい” is classified into Type 1, “〜〇”.

<table>
<thead>
<tr>
<th>Type</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Type. 1</td>
<td>〜〇</td>
</tr>
<tr>
<td>Type. 2</td>
<td>〜〇〜</td>
</tr>
<tr>
<td>Type. 3</td>
<td>〜〜〇</td>
</tr>
<tr>
<td>Type. 4</td>
<td>〇〜△</td>
</tr>
<tr>
<td>Type. 5</td>
<td>〜〇〜△</td>
</tr>
<tr>
<td>Type. 6</td>
<td>〜〇〜△〜□</td>
</tr>
<tr>
<td>Type. 7</td>
<td>〜〇〜△〜□〜</td>
</tr>
</tbody>
</table>

Figure 1. Structure Types of Sentence Patterns.

Each entry included in the Japanese sentence-pattern dictionary (Ask Shuppan, 2008) is first classified into one of the above seven structure-types, then segmented into multiple elements according to the number of “〜” and other symbols, and finally stored as a whole record into our digital dictionary. In the second step, the matching conditions to be examined for each “〜” have been rewritten to conform to Cabocha\(^4\), a free morphological analyzer, which is used to analyze input sentences later in Section 3. Our digital dictionary is composed of 371 records, each containing the Structure-type, the matching conditions as described above, and other necessary information as well. One of the authors has spent 1.5 months in editing this digital dictionary manually.

\(^4\)https://code.google.com/p/cabocha/
3. Sentence-pattern Examination

This section describes the flow of our system. After a Japanese sentence is composed, it is first segmented into multiple morphemes using Cabocha at the back. Then the pre-created digital dictionary is employed to examine whether the sentence is likely to contain a sentence pattern. If the answer is positive, another examination will be conducted to see whether the sentence pattern has been properly used. Finally, a feedback will be prompted to the user telling the examining result and a guide as well in case the usage is not correct. Figure 2 shows the algorithm for the specific case “〜○”.

Step1. Seek “○” in the input sentence.
Step2. Get the part-of-speech (POS) and conjugation information of “〜” using Cabocha if Step1 returns a success.
Step3. Compare the POS and the conjugation information of “〜” with those in the sentence-pattern dictionary for the corresponding “〜○”.
Step4. Prompt the user with a guide on how to use the sentence pattern properly in case a mismatch occurs in Step3.

Figure 2. Examining Steps for Sentence Pattern “〜○”

 Algorithms for examining other structure types in Figure 1 are similar to “〜○”, but a little more complicated. A screen shot of our system is shown as Figure 3.

Figure 3. A Screen Shot of the Interface of Our Web System

4. Experiments for Sentence-pattern Examination
We have conducted some experiments to examine the effectiveness of our approach. We employ two kinds of test dataset: D1 and D2. D1 is composed of 200 correct illustrative sentences extracted from a book on Japanese sentence patterns (ALC Shuppan 2007). Each illustrative sentence in D1 contains at least one sentence pattern. D2 contains 200 sentences which have been extracted randomly from a corpus (Tomoya Mizumoto etc., 2011) generated by using Lang-8, a language learning platform where native speakers correct what learners have composed\(^5\). Compared with D1, D2 is much closer to the practical input sentences. We hope to observe the difference arising from correctly used sentence patterns and the opposite by this means.

<table>
<thead>
<tr>
<th>Table 1: Experimental results for D1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of sentence patterns discovered</td>
</tr>
<tr>
<td>Number of true sentence patterns contained</td>
</tr>
<tr>
<td>Number of true sentence patterns discovered</td>
</tr>
<tr>
<td>Number of sentence patterns with correct feedbacks</td>
</tr>
<tr>
<td>Precision</td>
</tr>
<tr>
<td>0.93</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 2: Experimental results for D2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of sentence patterns discovered</td>
</tr>
<tr>
<td>Number of true sentence patterns contained</td>
</tr>
<tr>
<td>Number of true sentence patterns discovered</td>
</tr>
<tr>
<td>Number of sentence patterns with correct feedbacks</td>
</tr>
<tr>
<td>Precision</td>
</tr>
<tr>
<td>0.58</td>
</tr>
</tbody>
</table>

Table 1 and Table 2 show the experimental results on D1 and D2. Precision, Recall, and F-value are calculated to measure how effective our approach is in discovering sentence patterns. Feedback Precision indicates the success rate by which proper feedbacks have been given for correctly discovered sentence patterns. A significant performance degradation could be observed when the test dataset varies from correct sentences to real-world texts, whereas the latter is what we have to consider and important when evaluating learning support systems developed especially for less capable users.

<table>
<thead>
<tr>
<th>Table 3: New experimental results for D1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of sentence patterns discovered</td>
</tr>
<tr>
<td>Number of true sentence patterns contained</td>
</tr>
<tr>
<td>Number of true sentence patterns discovered</td>
</tr>
<tr>
<td>Number of sentence patterns with correct feedbacks</td>
</tr>
</tbody>
</table>

\(^5\) http://lang-8.com/
Table 4: New experimental results for D2

<p>| | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of sentence patterns discovered</td>
<td>230</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of true sentence patterns contained</td>
<td>180</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of true sentence patterns discovered</td>
<td>164</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of sentence patterns with correct feedbacks</td>
<td>154</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Based on the initial experimental results, we have modified Step1 in Figure 2 by incorporating morphological analysis. After locating “○” in the input sentence, we conduct a morphological analysis on the whole sentence to see whether “○” is segmented as a separate morpheme or partially attached to other morphemes around. Only sentences containing separate “○” are left over for further process. Table 3 and Table 4 show the new experimental results.

It is obvious that tightening the conditions for discovering sentence-patterns improved the system’s performance, especially for real-world texts. The Precision has increased from 0.58 to 0.71, indicating the effectiveness of the modification in algorithm. Totally, we have obtained a reasonable F-value for discovering sentence patterns and a satisfying Feedback Precision. We believe the system could be helpful for less capable users who are trying to use sentence patterns during the process of free composition. Among all the sentence patterns discovered by the system, about 30% might not really be sentence patterns according to the experimental results in Table 4. However, we don’t consider this as a major issue. Users could easily ignore the over-discovered sentence patterns in texts composed by themselves.

However, there are some drawbacks with this approach.

1. false analytical results from the morphological analysis
2. incompleteness of the sentence-pattern dictionary.
3. confusions between normal expressions and sentence patterns.

The first problem arises from the accuracy of the morphological analyzer we have been using. The second one is attributable to the exhaustiveness of the sentence-pattern dictionary we adopt. Some conditions to be examined for each “～” in a sentence pattern are not exhaustive. And this leads to some discovering failures as a result. We might be able to remove the these problems partially from our system by trying other analytical tools or sentence-pattern books. The last problem indicates another case where a normal expression is over-discovered as a sentence pattern. Here is an example.

**Input:**
ともだちのミシャールさんのうちにあそびに行きました。
(I hang out with Mishaal to her house.)

**Feedback:**
「～うちに」の文型を使っているようです。- OK
(Correct usage with the sentence pattern “～うちに”)
The system has recognized “うちに” as a sentence pattern meaning *before you know it / while*, and prompted a message showing that the user has used this sentence pattern correctly, whereas “うちに” could also be used to express *to someone’s home* which is correct here according to the context of this input sentence. This issue is hard to handle and we might need some extra statistical information to make better decisions.

5. Conclusion

We focus on sentence-patterns in Japanese and have developed a Web system to support sentence-pattern learning. Specifically, our system is able to discover a sentence pattern from an input Japanese sentence by using a digital sentence-pattern dictionary generated in advance. After that, our system determines whether the sentence pattern has been used properly and gives some instructions on how to use the sentence pattern correctly if the answer is negative. The difference between our system and Reading Tutor, an existing system also capable of discovering sentence patterns, is that our system does not expect correct input sentences at all, whereas Reading Tutor only accepts sentences containing sentence patterns that have been used correctly.

Experiments have been conducted to examine the effectiveness of our approach. On the whole, we have obtained a reasonable F-value for discovering sentence patterns and a satisfying Feedback Precision. We believe the system could be helpful for less capable users who are trying to use sentence patterns during the process of free composition.

Our future tasks include increasing the precision further in discovering sentence patterns, and a questionnaire by less-capable Japanese learners to examine the effectiveness of our Web System through practical application.

References

Abstract: Speech is one of the most natural ways of communication between human. In recent years, the spoken dialogue systems on human machine interaction (HMI) is more and more popular. In order to develop effective and natural human machine interaction, the corpus collected is relatively more important. Due to various types of corpus, classifying the corpus is a needed process. In this paper, we collected, transcribed, and classified the corpus, we named Chiayi Chinese Conversation Dialogue Corpus (CYCCDC). We collected this corpus with multiple ways, and then we arranged and classified this corpus. The corpus includes multiple useful information to research spoken dialogue system and human communication field. The CYCCDC includes tourism information, food information, clothing, housing information, traffic information and part of Orange Technology, such as elderly health care and accident handling. This corpus can be extensively applied, such as tourism plan in spoken dialogue system.

Keywords: CYCCDC, Chinese corpus, tourism, orange technology, spontaneous speech.

1. Introduction

With the advance of the internet and technology, people chat with each other not only through voice. People can also make a communication by text with various well-developed online instant messaging software, such as Skype, Line, Aim, and so on. Since text is easier to record than voice, it is helpful for searching history record or finding out particular conversation information with proper classification in the future. There is some corpus analysis as (Agrawal, 2011) which is classified the emotions of the Hindi corpus. In (Jia et al., 2011), the present study systematically states the construction of the corpus on the English learners in Asia. There is also some conversation corpus, which has been collected in (Bechet et al., 2012), the goal of this paper is to reduce the development cost of speech analytics systems by reducing the need for manual annotation. In (Takezawa et al., 2002), they collected the travel conversation corpus and a broad-coverage bilingual basic expression corpus, and they compared the characteristics of vocabulary and expressions between these two corpus. In this paper, we collect the conversation transcript documents and analyze the topic classification. Finally, we gave each conversation script a topic classification. In particular, the corpus of care and accident handling are classified into at the category of Orange Technology (Wang & Chen, 2011). Orange technology is the idea mentioned by the National Cheng Kung University professor Jhing-Fa Wang. The main idea of orange technology is to bring health, happiness and care for human. It also include elderly health care and child care. Besides, people can enhance safety and quality of humanism to foreign culture with orange technology. We also collected other corpus on several topics: tourism, food, sports, solicitude and others. Each topics can be classified into more categories. Ohtake et al. proposed a new corpus of consulting dialogues which is designed for training a dialogue manager (Ohtake et al., 2010). They also collected more than 150 hours of tourist guidance dialogue. In section 2, we describe the method how we collect corpus. Then we explain how we arrange and classify in section 3. In section 4, we show that our corpus can be applied in various fields. Eventually, we discuss the future work and conclusions in section 5 and 6.
2. Method

In this paper, we collected the Chiayi Chinese conversation dialogue corpus to increase the variability of response sentence for the spoken dialogue systems. Our approach is to ask at least two people for a chat and record their conversation. There are several rules as following:

- Sentence composed by the conversation dialogue (unlimited number of the sentences)
- Period of each conversation occurred at least 8 minutes
- Each dialogue included at least two topics
- Each conversation took by native Chinese speakers

The rule 1 is established to make the system spoken dialogue more humane. The purpose of the rule 2 is to let speakers converse in a spontaneous way. The rule 3 is helpful to increase the variability of contents that can enrich response sentence. In order to reduce Chinese grammar errors and obtain a native Chinese spoken dialogue, we augmented the rule 4. We obtained 392 audio files which length is at least 8 minutes, then turned the audio files into the transcribed text files. There are at least 20 sentences in each transcribed file. The sentences are related to the fields of tourism information, elderly care and accident disposal, which included health, unexpected events, food, traffic, housing, clothing and others. An example of the conversation is shown in Figure 1.

```
Speaker A: 這個周末我也要去嘉義玩
Speaker A: 希望也能有好天氣。
Speaker B: 你要去嘉義哪裡玩？
Speaker A: 先去看射日塔之後再去阿里山看日出吧。
Speaker B: 二二八紀念碑也可以去看看。
Speaker A: 要怎麼去呢？
Speaker B: 你可以坐公車去。
Speaker A: 附近有加油站嗎？
Speaker A: 我想開車去嘉義。
Speaker B: 附近過橋有一家中國石油。
Speaker A: 嘉義有甚麼好吃的在地小吃嗎？
Speaker B: 有一間簡單火雞肉飯不錯吃。
```

Figure 1. An example of the conversation.
First, the conversations must be recorded in a quiet room. Speakers need a computer, a microphone, and the recording software Praat before starting a conversation. Audio file format is designated as the WAV format, with a bit rate of 256 kbps and a sampling rate of 16 kHz. The sampling resolution in 16 Bit is recorded in a mono mode with the PCM. The speakers can think about what they want to talk about before recording. Each conversation should be taken at least 8 minutes and recorded at least 24 minutes in total. An example of Spectrogram is shown in figure 2.

The Chiayi Chinese conversation dialogue corpus has been recorded by 199 speakers, which are the students of National Chiayi University, Taiwan. There are 88 native Taiwan females and 111 native Taiwan males. The age range of these speakers is from 18 to 21. In total, there are 399 audio files and 27.5 hours of conversation speech. Consequently, the Chiayi Chinese conversation dialogue corpus comprised 9,734 sentences of 399 transcribed text files.

3. Corpus Analysis

<table>
<thead>
<tr>
<th>Topics</th>
<th>Instruction</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tourism</td>
<td>Dining Location Lodging Location Modes of Transport (Living Travel) Navigation Information (Modes of Transport) Traveling Spot Shopping Information (Living Travel) Cost of Time Weather/Climate Care Agency (Location)</td>
</tr>
<tr>
<td>Food</td>
<td>Dining Place (Restaurant, Snack Bar) Transportation (The Modes of Transport Go for Dining) Cost of Time (Meal Time)</td>
</tr>
<tr>
<td>Healthy Care</td>
<td>Transportation (Healthy Care) Navigation Information (Healthy Care) Shopping Information (Healthy Care) Weather/Climate (Health Care) Health Status</td>
</tr>
<tr>
<td>Accident Handle</td>
<td>Emergency Incident Detection (Fall Down, Faint, Asthma Attack)</td>
</tr>
<tr>
<td>Recreation</td>
<td>Leisure Entertainments (Knowledge/Reading, Recreational, Artistry, and so on)</td>
</tr>
<tr>
<td>Solicitude</td>
<td>Health State Greeting Consolation Family/Neighbor Trivia Working And so on</td>
</tr>
<tr>
<td>Sports</td>
<td>Ball Running Swimming Hiking And so on</td>
</tr>
<tr>
<td>Others</td>
<td>Not classified in the above topics</td>
</tr>
</tbody>
</table>

In this section, we describe how we design and classify these collected corpus. Due to a wide variety of corpus, we sorted all the corpus before the classification. It will make developers more convenient while using our corpus. And we will introduce main topics in section 3.1, then spoken speech phenomenon in section 3.2. In section 3.3 is describe about what is our method to distinguish topic in the sentences.
A “topic” is an abstract or a representative summary of the contents in dialogues. Developers can comprehend key points with the information of the topic instead of reading whole contents of dialogues. We can also classify conversations into appropriate categories. For example, a speaker says: “這個週末連假我們去嘉義的阿里山走走吧”, this sentence not only can be regarded as an effect of the topic “tourism” from a conversation, but also speculate from the keyword that the topic of this dialogue might be “tourism”. If there exists a topic in one interpersonal conversation this conversation would not be interrupted easily. Besides, this uninterrupted conversation can be continued to develop new contents or changed into different topic timely can motivate speakers for keeping the interpersonal conversation. Therefore, we designed 8 topics to classify the corpus. There are tourism, food, healthy care, accident handle, recreation, solicitude, sports and others. Each topic can be subdivided into a lot of classes as shown in table 1. According to transcription which is tagged each topic by speakers. There are 2,064 sentences in tourism, 1,042 sentences in food, 177 sentences in healthy care, 268 sentences in accident handle, 1,801 sentences in recreation, 2,128 sentences in solicitude, 513 sentences in sports and 1,741 sentences in others. The distributing graph is shown in figure 3. There are several issues need to be discussed.

3.1 Main Topics

According to human thought, the dialogue is nothing more than their food, clothing, housing, traffic and entertainment for main topic. For example, an old sick people want to travel, then he goes to some fun places, eat goodies and experience what he could not engage in leisure activities during illness. In this example, we designed the 7 topics, including the examples mentioned in the travel and food, while healthy Care, accident handle and solicitude are especially designed for the elderly.

3.2 Spoken speech Phenomenon

Through listening to the recorded speech, we found that corpus contains many voices speaking tone, because people thinking and emotion factors. Due to the recording of the conversation dialogue is not like reading speech which has a transcript. There is some spoken speech phenomenon. The most common of spoken speech phenomenon is particles, and the other is paralinguistic phenomena, the situation of pronunciation is not correct that is pronunciation error, and non-native language (Chang et al., 2005). There are an example of paralinguistic phenomena shown in figure 4. Because of the speakers conversed spontaneously, we classify those corpus to other topic.
3.3 Keywords of topic

For example, a speaker A says: “最近有聽說什麼好吃的美食嗎?”, a speaker B says: “有阿!聽說在嘉義奮起湖有好吃的鐵路便當”. There has different keyword of topics in this example. We will judged according to the context of the sentence belongs to what topics.

4. Applications

The most application of corpus is automatic speech recognition (Wu et al., 2014). There is also another application of corpus, such as generating responses sentences. Since the CYCCDC contains many different topics of sentences and rich vocabularies, the CYCCDC can applied in many aspects. For instance, developers can utilize the CYCCDC to generate responses sentences with some tour information for a tourism planning system, or refine an existing dialogue system with corpus of solicitude. The developers also can exploit the CYCCDC to analyze what do people talk about or when they want to have a travel. The CYCCDC is a conversational corpus so that the generated responses sentences of a dialogue system can make users feel like having a conversation with an actual human. This type of dialogue system is called as chat oriented dialogue system. Banchs et al. have also proposed the informal response interactive system (IRIS) (Banchs et al., 2012), which is a chat-oriented dialogue system based on the vector space model framework.

5. Future Work

In the future work, we will continue to improve the consistency of the database. With more corpus from the refined CYCCDC, all the related applications can be expected to be improved greatly, and can be applied in more studies. The spoken speech phenomenon also is still a research issue.

6. Conclusion

In this paper, our corpus was recorded and transcribed by the speakers. Different from other common corpus, the Chiayi Chinese conversation dialogue corpus is based on conversation for academic research but also substantial contribution. We designed 8 topics of the CYCCDC, which included the vast majority of tourism and solicitude. The native regional range of CYCCDC involved Yunlin, Chiayi and Tainan in Taiwan. Thus, the corpus also can apply in tourism planning system which
focused on these regions. Although all the sentences are still in the testing phase, but the quality of the sentences is adequate enough for doing researches. The other parts of corpus in CYCCDC are mainly collected for the Orange Technology, which focused on health care and accident handling for children and elders. The CYCCDC also involved some dialogue about disadvantaged groups or people who need social care. This part of the corpus can assists with the academic research of society. Certainly, the CYCCDC also contained some of common corpus, such as recreation, sports and other categories that are irrelevant to the main collection. Because the CYCCDC is a real conversation between humans, the CYCCDC is also helpful for a chat-oriented dialogue system.

Acknowledgments

This work is supported in part by the National Science Council, Taiwan, R.O.C., under the project grant numbers NSC 102-2221-E-415-006-MY3.

References


Partial and Synchronized Caption Generation to Develop Second Language Listening Skill

Maryam Sadat MIRZAEI *, Yuya AKITA a, Tatsuya KAWAHARA a
a Graduate School of Informatics, Kyoto University, Japan
*maryam@ar.media.kyoto-u.ac.jp

Abstract: Captioning is widely used by second language learners as an assistive tool for listening. However, the use of captions often leads to word-by-word decoding and over-reliance on reading skill rather than improving listening skill. With the purpose of encouraging the learners to listen to the audio instead of merely reading the text, the study introduces a novel technique of captioning, partial and synchronized, as an alternative listening tool for language learners. Using TED talks as a medium for training listening skill, the system employs the ASR technology to synchronize the text to the speech. Then, the system uses the learner’s proficiency level to generate partial captions based on three features that impair comprehension: speech rate, word frequency and specificity. To evaluate the system, the performance of Kyoto University students in two CALL classes was assessed by a listening comprehension test on TED talks under three conditions: no caption, full caption and the partial-and-synchronized caption. Results revealed that while reducing the textual density of captions to less than 30%, the proposed method realizes comprehension performance as well as full caption condition. Besides, it performs better than other conditions on new segments of the video without captions.

Keywords: Computer-Assisted Language Learning, Automatic Speech Recognition, Listening Comprehension, Word Frequency, Speech Rate

1. Introduction

The process of learning a foreign language involves mastering different skills such as listening, speaking, reading and writing. Of these, acquiring listening often entails a complex cognitive process and demands the use of different strategies which in turn make a phase of frustration for many language learners (Leveridge and Yang, 2013). In order to improve listening, one must be exposed to authentic and comprehensible input. Authentic input, however, makes listening more challenging especially when the phonological systems of the first and the second language are distant (e.g. Japanese vs. English).

Listeners can overcome this problem by benefiting from assistive tools such as “captioning” that textualizes the verbatim speech and makes it more recognizable through neatly dividing the word boundaries. Nevertheless, when it comes to using captions, both language learners and teachers face a dilemma. In fact, when reading captions is part of watching a video, learners often rely on their reading skill to compensate for their listening skill deficiencies, whereas in a real-world communication, learners should solely use their listening skill as no assistive tools are available.

To address these issues, this study proposes a new method of captioning, “partial and synchronized” as an alternative tool for enhancing second language (L2) learners’ listening comprehension skills. The term “synchronized” captioning is to present caption text word by word aligned in precise timing with the speech signal of the respective words, which effectively shows the correspondence between words and the audio channel. This method is realized by the automatic word-level alignment feature of automatic speech recognition (ASR) technology, which precisely maps each word to its corresponding speech signal. In the “partial” captioning method we select a subset of words from the transcript and present them in the caption while hiding the rest of the words. Although seems similar to keyword captioning, in this method “important” words are not the selection criteria. Instead, words that impair comprehension or the ones beyond the learner’s current level of competence form the basis of this selection. Moreover, the selection of keywords is content-specific
and does not consider the proficiency level of the learners, whereas the features of the proposed method are tuned to the learner’s knowledge to meet the requirements of each individual.

Unlike conventional captions, in Partial and Synchronized Captioning, comprehension cannot be gained by solely reading the captions, but by listening to the audio and reading only for difficult or unrecognizable words. Thus the method is effective for reducing learners’ dependence on captions.

Following this introduction, this paper reviews previous studies and describes the proposed technique of captioning. Then, the experimental procedure together with the results is demonstrated and a discussion over the findings is addressed. This paper ends with conclusion and future directions.

2. Literature Review

2.1 Captioning and L2 Listening Comprehension

To overcome the listening problems, assistive materials, such as captions, are used to help L2 listeners. Captioning is defined as “visual text delivered via multimedia that matches the target language auditory signal verbatim” (Leveridge and Yang, 2013, p.1). Captions neatly demonstrate the word boundaries without being affected by accent, pronunciation and audio deficiencies (Vanderplank, 1993) and allow the learners to parse the speech stream into meaningful chunks, an essential process for learning (Ellis, 2003). A considerable amount of literature has been published on various beneficial effects of captions. Some of these studies have investigated the effect of captioning on vocabulary acquisition (Bird and Williams, 2002; Griffin and Dumestre, 1992), reading development (Bean and Wilson, 1989), word recognition (Bird and Williams, 2002; Markham, 1999) and listening comprehension (Danan, 2004; Garza, 1991; Markham, 1999; Montero Perez et al., 2014; Vanderplank, 1993; Winke et al., 2010).

For instance, Garza (1991) conducted an experiment with 70 high-intermediate learners of English and 40 three to four year learners of Russian, and assessed their comprehension of videos with/without captions. His results indicated significant improvement on the captioning condition in both groups. Studies in Japan such as Suzuki (1996) reported the positive effect of English caption on Japanese listening comprehension development.

The type and manner of captioning may influence the effect of this assistive tool on language learning. Garza (1991) suggests using various types of open captioning, such as verbatim, paraphrase and keywords as means of training listening skill.

2.2 Aligned and Synchronized Captioning

Correspondence between caption and speech may also affect the learning process. Advancement of speech technology has enabled precise text-to-speech alignment. Munteanu et al. (2007) used ASR to generate transcripts of webcast lectures for examining native speakers’ comprehension on the videos. They found out that ASR generated transcripts are useful when word error rate (WER) is lower than 20%. This finding was generalized to L2 learners in a study by Shimogori et al. (2010) who suggest that captions with 80% accuracy improve the understanding of Japanese learners of English.

Accordingly, “karaoke-style” display, where the text is highlighted in colors as the audio moves by, has gained some instructional value. Bailly and Barbour (2011) developed a system that exploits the alignment of text with audio at various levels (letters, phones, syllables, etc.). This system uses a data driven phonetizer trained on an aligned lexicon of 200,000 French entries to display a time-aligned text with speech at phoneme level. The results showed that the multimodality of synchronous reading systems is beneficial for overcoming the problem of word decoding in a text/audio-only environment.

It should be noted that this method may lead to over-reliance on the caption and needs to be refined. This can be accomplished through highlighting only particular words or sentences in the caption, as in keyword captioning.

2.3 Keyword Captioning
Guillory (1998, p.95) examined the use of keyword captioning for learners of French. The results demonstrated that students who received keyword captions performed as well as those who received full captions. Guillory discussed that “learners no longer need to be subjected to a volume of text to read; they can in fact comprehend authentic video with considerably less pedagogical support”.

In a recent study by Montero Perez et al. (2014), the perceived effectiveness of keyword captioning is criticized. The study investigated the effect of full text captions and keyword captions versus no captioned condition. The results demonstrated that full captioning group outperformed the other two groups on the global comprehension questions while both the keyword captioning and the no-captioning group had equal performance on this test. Analysis of the responses received from the keyword-captioning group revealed that this type of captioning is distractive. According to the researchers, a plausible explanation may be the salient and irregular appearance of the keywords on the screen, which causes distraction. However, not every learner can benefit from presenting the keywords in captions since the selection of keywords is content-specific and may not provide each learner with his/her required amount of support. In line with this assumption, Guillory (1998) noted that the keyword captions used for her study contained a tiny portion of the total script, which may not have provided enough information for the beginners.

2.4 Limitations on Captioning

In spite of the beneficial aspects of captioning, there are some criticisms on the use of this assistive tool. It is skeptical whether learners provided with captions are training their listening or their reading skills. Kikuchi (1995) examining subtitles in Japanese and captions in English reported that students who watched the movie with Japanese subtitles merely read the text without listening to the movie. Using an eye tracker, Winke et al. (2013) investigated learners’ use of captions and reported that learners read the captions on average 68% of the time it is on the screen.

On one hand, the learner needs to be able to deal with real-world situation where there is no access to any supportive tool, and on the other hand we cannot expect a non-native listener to follow the authentic input without any support. Hence, the listening instruction should focus first and foremost on assisting the language learners to cope with aural input difficulties while maintaining a tendency to develop compensatory strategies for listening in real-time. Thus, further research should be conducted to investigate an effective method for assisting learners to gain adequate comprehension, without becoming too much dependent on captions.

3. Proposed Method: Partial and Synchronized Captioning

We propose a new type of captioning called Partial and Synchronized Captioning (hereinafter, PSC). In this method the text is synchronized to the speech in word-level and only a subset of words are shown in the caption while the rest are masked to keep the learner listening to the speech. Thus, this method consists of two components; synchronization and partialization where the two are complementary and counteract the demerits of one another.

First, synchronized caption is automatically generated; word-level synchronization of text with speech is realized by ASR. The word-level alignment, which synchs each word with the speaker’s utterance, presents the phonological visualization of the words and thus leads to improvement in aural word recognition skills through mapping between the speech stream and the verbatim text.

Moreover, this method neatly presents word boundaries, which often cannot be easily recognized in authentic speech input. Synchronized captions, although in favor of many language learners, may bring too much assistance for the learner and makes them more and more dependent on the caption (Vandergrift, 2004; Garza, 1991). In order to solve the disadvantages of this method, we propose partial captioning which builds on synchronized captions to provide the students with reduced transcription of the videos in order to better train them for real-world situations.

This method can act as an intermediary stage before the learner is totally independent of captions. In this method, the filtering process of words to be presented takes into account not only the hindering factors of comprehension, but also the assessed knowledge of the learner. Hence, adjusted
to a particular learner’s need, the method selects words which are beyond the proficiency level of the learner. However, if using partial captions alone, as in keyword captioning, the students are often distracted by the sudden and irregular appearance of a word on the screen (Montero Perez et al., 2014). Nevertheless, this problem is mitigated by synchronization in PSC.

To conclude, this new tool, PSC, is anticipated to make the learner less dependent on caption and more prepared to handle listening in real-world situations. Table 1 summarizes the advantages of PSC compared to other captioning methods and Figure 1 shows the screenshot of a generated PSC.

Table 1: Comparison of different caption methods.

<table>
<thead>
<tr>
<th>Caption Type</th>
<th>Advantage</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Full Caption</td>
</tr>
<tr>
<td>Aid word boundary detection</td>
<td>✓</td>
</tr>
<tr>
<td>Speech-to-text mapping</td>
<td>✓</td>
</tr>
<tr>
<td>Avoid over-reliance on reading</td>
<td>✓</td>
</tr>
<tr>
<td>Avoid being distractive</td>
<td>✓</td>
</tr>
<tr>
<td>Automatic</td>
<td>✓</td>
</tr>
<tr>
<td>Adjustable to learners’ knowledge</td>
<td>✓</td>
</tr>
<tr>
<td>Adjustable to the content</td>
<td>✓</td>
</tr>
</tbody>
</table>

3.1 Feature Selection

In order to decide which words to show in the caption and which ones to hide, the following features were picked as the selection criteria. These features were chosen for being identified as major contributing factors in listening comprehension impair. Besides, these factors can be quantified automatically and are easy to be implemented.

3.1.1 Speech Rate

Previous studies showed that high speech rate can negatively affect L2 listeners’ comprehension (Dunkel, 1994) and this is even true for native speakers (Wingfield et al., 1985). For Japanese learners of English, particularly, fast rates of speech and inability to perceive the sounds in English are the major factors to impair comprehension (Osuka, 2007). Some studies suggested modification of speech rate as a solution, however, this is not close to real-world situation. Instead, we provide the learner with PSC that presents words/phrases uttered faster than normal rate of speech, or that of tolerable for the learner.

3.1.2 Word Frequency

When the vocabulary chosen by the speaker exceeds the vocabulary size of the listener, comprehension will be impeded. In such cases the unknown words confine the learner’s attention, and as the speech proceeds the learner cannot pursue the subsequent parts. In other words, the listener invests a lot of time trying to understand what s/he missed (Goh, 2000).

The frequency of word usage in a language is a measure to assess word difficulty. For instance, learners are less likely to be familiar with low-frequency words (Nissan, 1996). Word frequency is calculated based on its occurrence in spoken or written corpora. A well-cited paper by Nation (2006) categorizes English vocabulary into High-frequency (the most frequent 2000-3000 word families), Mid-frequency (anything between 3000-9000 word families), and Low-frequency (beyond the 9000 frequency band). The term word family here refers to a base word and all its derived and inflected forms that can be recognized by a learner without having to learn each form separately.
To assist L2 listeners, PSC presents words or phrases, which are less frequent and hence make comprehension difficult.

### 3.1.3 Word Specificity

The occurrence of specific words in a video would make comprehension difficult since limited knowledge of academic words is often seen as a reason for L2 listening comprehension deficiency (Goh, 2000). Thus, when considering word frequency, it is important to consider word specificity as well. Using academic talks as the material for this study, this feature is also taken into account in PSC.

### 4. System Architecture

Figure 2 depicts the data flow and main components of the system. The procedure of generating a PSC starts with an alignment phase where the ASR system outputs the transcript with estimated word timing, which is aligned and adjusted with the given transcript of the caption. Next, word frequency, word specificity and speech rate are used to serve as the selection criteria for making PSC. The feature extraction module further processes the transcript and converts it into a feature vector for the decision making module.

A rule engine in the decision making module decides whether a word should be shown or not. This decision not only depends on the features, but also relies on the user input (i.e. quiz results).

In the formatting and display module, the captions are altered as the desired output of the system. Being synchronized with the utterance of the word, the corresponding dictation of the word (or character mask) should appear on the screen. Eventually this module plays back the media with the generated caption, and offers a pre-made comprehension test afterwards.

**Figure 2.** Data flow and the main components of the system.

**Figure 3.** Percentage of words shown in PSC for intermediate learners

#### 4.1 Alignment Module

The input data is composed of a video and its transcript text. To obtain the time tag of the tokenized words automatically, the audio should be ripped from the video to be passed to our ASR system, Julius v4.3.1 (Lee and Kawahara, 2009). Since Julius itself is a language-independent decoding program, it is possible to make a recognizer of a language, given an appropriate language model and acoustic model for the target language. The performance of ASR largely depends on these models. In this study TED talks were selected as the material. Thus, for precise alignment to take place, it is necessary to train the ASR models using a matched corpus, in this case TED talks. This model training was done in our laboratory, based on the lightly-supervised training approach using 780 TED talks (Naptali and Kawahara, 2012). The transcript and ASR output then got aligned using the force-alignment procedure.
4.2 Feature Extraction Module

This module extracts the main features and calculates them. The following elaborates on these features.

4.2.1 Speech Rate

The speech rate is often measured in Words per Minute (WPM) or Syllables per Second (SPS). The former may be affected by pauses and change of speech rate within a minute which causes inaccurate measurement while the latter is more suitable to measure short speeches and thus is used in this study.

The first step to calculate this feature is to estimate the speech rate where we need to count the number of syllables in each word, and then calculate the duration of its utterance. Calculation of the syllables is based on the structural syllabification of the corresponding text, which was realized using Natural Language Toolkit (NLTK). The full calculation of speech rate requires the duration of a word, which is calculated by the time tags obtained in the alignment phase after excluding the long pauses.

4.2.2 Word Frequency

Word frequency is defined by referring to corpus-based studies. Nation (2006) has designed 25 word family lists each including 1000 word families, plus four additional lists: (i) an ever-growing list of proper names; (ii) a list of marginal words including swear words and exclamations; (iii) a list of transparent compounds; and (iv) a list of abbreviations. The first two lists are carefully hand-selected while the rest are based on the following two famous corpora.

- The British National Corpus (BNC) which involves 100 million word collections of samples of written and spoken language from British English.
- Corpus of Contemporary American English (COCA), gathered by Mark Davies (from 1990 to 2012), includes 450+ million words. The corpus is equally divided among spoken, fiction, popular magazines, newspapers, and academic texts.

This study is based on aforementioned word family lists and COCA. Every word is lemmatized first, and the result is looked up for the word family, created offline from the COCA and BNC corpus. The family of the lemmatized word serves as the difficulty index. The word is also cross-checked with the spoken genre section of COCA.

4.2.3 Word Specificity

In this method specific words are determined using a popular catalogue called Academic Word List (AWL) by Coxhead (2000) which includes 570 headwords and about 3000 of academic words altogether. Besides, these words are cross-referenced with COCA’s academic words (Gardner and Davies, 2013) for more accuracy. The system is also capable of handling other features such as abbreviations, proper names, numbers, transparent compounds, and repeated appearance of words.

4.3 Decision Making Module

Based on the features, the system decides whether a word should be included in the final partial caption or not. This decision not only relies on the value of the features, but also considers general features.

In the first stage, the main features - word frequency, speech rate, and specificity - are accounted. If only one of them require a word to be shown, the word is marked to appear in caption. To decide on the word frequency feature, a vocabulary size test (Nation and Beglar, 2007) is employed to assess the vocabulary size of the learner and to determine the appropriate frequency threshold for him/her. Similarly, a decision about whether a word should be a candidate for being shown in partial citation is taken by comparing the calculated speech rate of the word to that of preferable for the learner. Thus, if the utterance of the word (measured by speech rate feature) is faster...
than the tolerable threshold of the learner, the word will be shown in caption as a textual clue. This threshold can be adjusted by the user.

In the second stage, the general features act on each word. The features are either excitatory or inhibitory. The decision on general features is made on top of the first stage. For instance, abbreviations and proper names are being marked to be displayed while interjections are marked to be discarded.

The third stage of decision-making is about the sequence of the words that should be readable for the learners. The rules also handle words after numbers and words after “apostrophe s”.

4.4 Formatting and Display Module

This module generates the final partial and synchronized caption using the user display parameters. If the word is decided to be shown, it will be copied intact in the partial caption; otherwise a character mask (here we use “dots”) replaces every letter of the word. This will emulate the speech flow, by showing each and every word in the given speech in sync with their utterance. (e.g. “express” will be replaced by “........” and “don’t” will be replaced by “....”).

5. Experiment

Given the novelty of partial and synchronized captioning method, experiments were needed to evaluate the effectiveness of this technique. Thus, the study investigates the following questions:

- Do captioned videos result in better comprehension of video compared to non-captioned ones?
- Can the proposed captioning method substitute the conventional full-text captioning?
- Do proficiency differences affect the benefits obtained from the proposed captioning method?
- Does the proposed method help the learner comprehend the video better without any captions?

5.1 Participants

The participants were 28 and 30 Japanese students at Kyoto University ranging from 19 to 22 years old. These students were undergraduates of different majors who enrolled at a CALL course. The experiments were carried out over this course, in two different classes, for three consecutive sessions.

5.2 Material

Videos: The video materials of this research were selected from TED website which provides us with authentic videos plus almost accurate captions without the copyright issue (www.TED.com). The selection criteria were bound to “popularity” and “recentness” of the videos. The selection was carefully done to include only videos of native American speakers, to avoid the influence of other accents. All videos were trimmed to 5-minute meaningful segments.

Pre-study Vocabulary Size Test: A vocabulary size test created by Nation (2007) was used to evaluate the vocabulary reservoir of each student. The results of this test were used both as a placement criteria of dividing students into groups of proficiency and as a value to determine the frequency threshold for our caption generator. This test consists of 140 multiple-choice questions, with 10 items from each 1000 word family level. Since the caption generator uses the same word family lists as its references, the result of the test is appropriate to be set as our threshold.

Partial and Synchronized Caption Statistics: Taking into account the result of the vocabulary size test and the tolerable rate of speech, the system generates appropriate captions for learners with different proficiency levels. The percentage of words to be shown in the final caption does not exceed 30% for any of the videos as illustrated in Figure 3. This figure presents how the generated captions show fairly equal amount of words per video for a particular intermediate learner.

Comprehension Tests: After watching each video with assigned caption, the students were asked to take a listening comprehension test in the form of multiple choice and cloze test on summary.
5.3 Procedure

The study was conducted in CALL classes where students were provided with a 20 inch-wide screen and a headphone. Although the experiment was held in two different classes, the same procedure was adopted for both. Same videos were captioned with a different method (PSC↔FC) for each class.

We considered learner’s proficiency as a blocking factor, with three levels: “beginner”, “pre-intermediate” and “intermediate”. For the purpose of dividing the students into these three groups, the assessed vocabulary size together with the students’ TOEIC/CASEC scores were considered.

Each video, regardless of the caption type assigned to that was divided into two segments; 70% from the beginning and the rest of 30%. The students watched the first part of the video (70%) under one of these three conditions: no-caption (NC), full-caption (FC) and partial and synchronized caption (PSC). This was followed by a listening comprehension test. Next, the subjects were asked to watch the rest of the same video (30%) “without any caption” (regardless of the type of caption in the previous phase), and took another test. The procedure remained the same for all videos, while the type of caption was changed. To be more specific, the second part of each video is dedicated to evaluate students’ performance on a non-captioned video as in real-world condition.

6. Results and Discussion

The scoring system was easily constructed because of the objective format of multiple-choice and cloze-on summary items. One point was awarded for each correct answer to multiple-choice questions while partial credit (0.25) was given to each item in cloze test. The total score was finally calculated in percentage for all participants in each group. One-way ANOVA test was used to analyze the result of the tests and to investigate whether any statistically significant difference is found between different conditions under which the learners watched the videos.

Table 2: Comprehension performance of both classes on the first part of video with (NC, PSC, FC)

<table>
<thead>
<tr>
<th>Proficiency Level</th>
<th>Mean</th>
<th>SD</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Beg.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NC</td>
<td>28.7</td>
<td>13.6</td>
<td>19</td>
</tr>
<tr>
<td></td>
<td>34.7</td>
<td>11.8</td>
<td>19</td>
</tr>
<tr>
<td></td>
<td>43.3</td>
<td>15.1</td>
<td>20</td>
</tr>
<tr>
<td>Total</td>
<td>35.7</td>
<td>14.7</td>
<td>58</td>
</tr>
<tr>
<td></td>
<td>Beg.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PSC</td>
<td>42.0</td>
<td>16.7</td>
<td>19</td>
</tr>
<tr>
<td></td>
<td>52.0</td>
<td>17.5</td>
<td>19</td>
</tr>
<tr>
<td></td>
<td>64.0</td>
<td>18.0</td>
<td>20</td>
</tr>
<tr>
<td>Total</td>
<td>52.9</td>
<td>19.4</td>
<td>58</td>
</tr>
<tr>
<td></td>
<td>Beg.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FC</td>
<td>41.1</td>
<td>12.3</td>
<td>19</td>
</tr>
<tr>
<td></td>
<td>57.2</td>
<td>14.8</td>
<td>19</td>
</tr>
<tr>
<td></td>
<td>63.9</td>
<td>16.4</td>
<td>20</td>
</tr>
<tr>
<td>Total</td>
<td>54.2</td>
<td>17.3</td>
<td>58</td>
</tr>
</tbody>
</table>

* NC: No Caption  PSC: Partial and Synchronized Caption  FC: Full Caption

As shown in Table 2, analysis of the first part of the experiment (watching 70% of the videos) revealed a significant difference between NC (M = 35.7, SD = 14.7) condition and PSC (M = 52.9, SD = 19.4) and also FC condition (M = 54.2, SD = 17.3) at the p < .05. The results provide a positive answer to our first research questions, which concerns the effectiveness of PSC as compared to NC. However, no significant difference was found between the scores gained under PSC and FC conditions in this part of the experiment [F (1, 57) = 25, p = .62]. The findings suggest that PSC leads to the same level of comprehension as FC while providing less than 30% of the transcript.
Consequently, the assumption of our second research question is plausible and hence PSC can be used as an alternative to full captioning method for training L2 listening. Furthermore, the results reveal that students with different proficiency levels gained almost equal scores under PSC and FC conditions and could benefit from our method. A possible explanation for deriving such results may lie in the adaptability of PSC that considers the proficiency level of the learners for generating appropriate amount of caption for them and provides adequate assistance for any learner.

Table 3 presents the results of comprehension tests on the second part of the experiment where students watched the rest of videos without any captions immediately after having watched the first parts under different conditions (NC, FC, PSC).

In the second part of the experiment (30% without caption), the best performance is associated with the condition in which the learners first watched the video with PSC \[ F (2,118) = 20.5, \ p < .05 \] as compared to FC and NC. The findings highlight the effectiveness of PSC on preparing the learner for real-world situation where captioning is not available. While this result indicate a short-term enhancement partly because of adaptation to the video, this finding is still of value.

### Table 3: Comprehension performance of both classes on the second part of video without caption

<table>
<thead>
<tr>
<th>Proficiency Level</th>
<th>Mean</th>
<th>SD</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>NC</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Beg.</td>
<td>33.0</td>
<td>16.0</td>
<td>19</td>
</tr>
<tr>
<td>Pre. Int.</td>
<td>37.4</td>
<td>16.6</td>
<td>19</td>
</tr>
<tr>
<td>Int.</td>
<td>50.0</td>
<td>15.6</td>
<td>20</td>
</tr>
<tr>
<td>Total</td>
<td>40.1</td>
<td>17.4</td>
<td>58</td>
</tr>
<tr>
<td><strong>PSC</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Beg.</td>
<td>49.6</td>
<td>15.8</td>
<td>19</td>
</tr>
<tr>
<td>Pre. Int.</td>
<td>57.7</td>
<td>17.2</td>
<td>19</td>
</tr>
<tr>
<td>Int.</td>
<td>62.5</td>
<td>17.4</td>
<td>20</td>
</tr>
<tr>
<td>Total</td>
<td>56.6</td>
<td>17.3</td>
<td>58</td>
</tr>
<tr>
<td><strong>FC</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Beg.</td>
<td>38.3</td>
<td>13.5</td>
<td>19</td>
</tr>
<tr>
<td>Pre. Int.</td>
<td>40.4</td>
<td>11.9</td>
<td>19</td>
</tr>
<tr>
<td>Int.</td>
<td>49.3</td>
<td>12.7</td>
<td>20</td>
</tr>
<tr>
<td>Total</td>
<td>42.7</td>
<td>13.4</td>
<td>58</td>
</tr>
</tbody>
</table>

*NC*: No Caption  \( \text{PSC: Partial and Synchronized Caption} \quad \text{FC: Full Caption} \)

### 7. Conclusion and Future Work

The study introduced a novel technique of captioning, partial and synchronized, which is based upon speech rate, word frequency and specificity, to generate a smart type of caption that deals with limitation of previous methods. This method is based on the premise that the presence of infrequent or specific words and fast delivery of speech by the speaker hinder learner’s listening comprehension. Additionally, by synchronization, the system emulates the speech flow which facilitates text-to-speech mapping and avoids the salient appearance of the words on the screen. Besides, to generate a suitable caption for a particular learner, the system assesses the tolerable rate of speech and vocabulary size of the learner and prepares the captions in accordance to his/her level of competence.

Evaluated in two CALL classes, the results of the experiment showed that students’ scores using PSC overtook that of the no-caption condition while resulted in almost equal comprehension as the full-caption condition. Furthermore, learner’s scores on a new segment of the video without caption was significantly higher than other conditions when they watched the video with PSC first. The finding highlights the positive effect of PSC in preparing learners for listening in real-world situations.

The results also indicate that our method can assist learners to obtain adequate comprehension of the video by presenting less than 30% of the transcript to them. Such a method is assumed to be
effective particularly for Japanese students who heavily rely on caption text in order to comprehend the content of the video. The findings further suggest that this form of captioning can be effectively incorporated into CALL systems as an alternative method to enhance L2 listening comprehension.

Long-term study requires both time and dedicated resources such as CALL classes that in this stage of the study was infeasible. Instead, we considered the immediate effect of the proposed method presuming a real-world situation by checking the learner’s comprehension of a new segment of the video without any caption after being exposed to our proposed method. Although the findings has shown comprehension improvement on a short-time adaptation experiment, given the nature of listening skill, overall improvement could not be realized unless the participants undertake long-term experiments, hence such an experiment is suggested.

References


Challenges in the annotation of article errors in Spanish learner texts

María del Pilar VALVERDE IBAÑEZ
*Department of European Studies, Aichi Prefectural University, Japan
valverde.mp@gmail.com

Abstract: Annotating learner texts with article error information is a difficult task. To identify which are the main difficulties for annotators, we carry out an annotation experiment in Spanish texts written by Japanese learners. Two expert and two non-experts raters annotate 300 noun phrases containing a definite, indefinite or zero article. We calculate inter-annotator agreement and analyse the sources of disagreement. We find article usage governed by pragmatic factors causes disagreement the most, while lexico-semantic factors are the most reliable. Finally, a learner corpus sample of 30,000 words is annotated with the revised annotation scheme.

Keywords: articles, error annotation, learner Spanish, inter-annotator agreement, reliability, corpus annotation

1. Introduction

The annotation of learner texts with error information is necessary for linguistic research as well as for the development of educational applications for language learning. While research has focused on the development of learner corpora and tools for English as a foreign language, the field of Grammatical Error Detection (GEC) is expanding and there is a need to develop resources for other languages. However, annotation and evaluation best practices are still an open issue.

First, inter-rater reliability for error annotation can vary widely: while for some errors, like number and gender agreement, rules are clearly defined, and using one rater may be acceptable, other kind of errors like article or preposition presence and choice are harder to annotate (Tetreault et al. 2010), so using only one annotator is not enough reliable. For article and noun number selection, for example, in Lee et al. (2009) raters found more than one valid construction for more than 18% of noun phrases. For prepositions, Tetreault and Chodorow (2008) found that even in native texts, “native raters can disagree with each other by 25% in the task of preposition selection”. In spite of this, learner corpora are typically annotated only once because double annotation would be too expensive, and few annotation projects provide measures of inter-annotator agreement (Rozovskaya and Roth, 2010; Lee et al., 2012). The need to improve annotation quality has been put forward by the NLP community, that has found difficulties for evaluating error detection systems in the last GEC shared tasks: in the first Helping Our Own task (2011), systems were penalized for valid corrections not annotated in the data while in last three tasks (H00 2012, CoNLL 2013 and 2014) teams could request the organizers to make changes in the annotation (Tetreault et al. 2014).

Second, as noted by Reidsma and Carletta (2008), there are different types of disagreement: chance disagreement, caused by random slips or lack of knowledge of the annotators, and systematic disagreement, due to different intuitions of the annotators or to a misinterpretation of the annotation guidelines. This distinction is crucial for the development of gold standards, since systematic disagreement can have a worse effect on machine learning than noise-like disagreement.

In this scenario, how can we improve the quality of annotations? Multiple annotations by more than one trained annotator is unrealistic for large projects and while crowdsourcing has shown good results for English preposition error annotation in a pilot study (Tetreault et al. 2010), more research is
needed to deal with other languages and error types. To date, there has not been found a good method to improve the quality and number of annotations for learner texts.

The goal of this paper is to investigate the main difficulties faced in the annotation of article errors in Spanish learner texts, so that measures can be taken to improve the quality of future annotation efforts. To do that, we carry out an experiment on article error annotation with a preliminary annotation scheme. In section 2 we describe the annotation principles, in 3 we briefly describe the data collection and annotation procedure, and in section 4 we investigate the sources of disagreement and main difficulties. In section 5 we apply the revised annotation scheme to a learner corpus sample and in 5 we present the conclusions.

2. Annotation principles

2.1. Article errors

In Spanish, articles can be definite (as in English *the*) or indefinite (in English *a/an*), and their form changes according to the gender and number of the noun they complement, as shown in Table 1 (base form in bold face).

Table 1: Spanish articles (gender and number)

<table>
<thead>
<tr>
<th></th>
<th>Definite</th>
<th>Indefinite</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Masculine</td>
<td>Feminine</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(\text{la})</td>
</tr>
<tr>
<td>Singular</td>
<td>\text{el}</td>
<td>\text{la}</td>
</tr>
<tr>
<td>Plural</td>
<td>\text{los}</td>
<td>\text{los}</td>
</tr>
</tbody>
</table>

The definite article *el* ‘the’ is the most frequent word in Spanish and article usage is also one of the most frequent grammatical errors\(^1\) among learners (specially for speakers of languages that do not have articles like Chinese, Japanese, Korean or Russian.), because article choice it is the result of interacting pragmatic, semantic, syntactic and lexical constraints.

We consider the following type of errors: missing article, extraneous article and article confusion. We are only concerned with article presence and choice, so we did not tag malformation (e.g. spelling or agreement errors) or order errors. A missing article occurs when the learner does not use any article but the noun phrase should contain one. An extraneous article occurs when the article used by the learner is not necessary (zero article is correct). A confusion error occurs when the learner used a definite instead of an indefinite, or vice versa, or when the learner uses a different type of determiner instead of an article.

2.2. Level of confidence in the judgments

It was expected that the annotators would sometimes be unsure about the acceptability of article usage in a given sentence, or unable to determine the most likely correction. With regard to the level of confidence in the annotators’ judgments, annotated corpora do not explicitly provide confidence levels for every annotated item. Only in some annotation experiments the annotators are asked to indicate their level of confidence (as “low” or “high”) (Tetreault and Chodorow, 2008).

We did not want to force the annotators to make a best guess in “difficult” sentences because that would lower inter-annotator agreement. Instead, we gave the possibility of marking such sentences as “difficult to judge” (as in Han et al. 2006), so later we could look at the sentences marked as problematic, and analyse what they have in common.

\(^1\) Fernández (1997) found 2.2 article errors per 100 words in a 4,433 words sample.
2.3. Number of tags

With regard to the number of possible analysis a sentence can receive, error-annotated learner corpora typically contain only one tag per error. However, the "single correct construction" approach has been questioned and in recent annotation efforts there is a tendency to allow the inclusion of several alternative codes for the same item (Lüdeling et al., 2005; Boyd, 2010; Lee et al., 2012; Rozovskaya and Roth, 2010). However, it is unattainable to list all possible interpretations for every error, so this is done only when the error analysis is doubtful. In our experiment, we decided to allow only one tag per item to detect the sources of disagreement. After the revision of the annotation scheme, double tagging would be allowed in some specific cases (4.3.2).

3. Experiment

We carried out an experiment on article error annotation with the following objectives. First, calculate inter-annotator agreement for this task, which can be considered as the limit for an automatic article error detection system. Second, analyse the types and sources of disagreement, to find out which are the main difficulties the annotators face when annotating article errors in learner texts, so that measures can be taken to refine the annotation scheme and future annotation can be improved.

3.1. Data collection

A teacher of Spanish as a Foreign Language extracted sentences containing at least one article error from students’ written assignments, 250 sentences for each kind of article (definite, indefinite and zero article). The same number of sentences, but with at least one correct article usage, was then collected from the same texts. The distribution of the data is as Table 2 shows. In every sentence only one highlighted noun phrase had to be annotated, so the number of sentences and the number of annotated noun phrases is the same.

<table>
<thead>
<tr>
<th></th>
<th>Definite</th>
<th>Indefinite</th>
<th>0 article</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Correct</td>
<td>50</td>
<td>50</td>
<td>50</td>
<td>150</td>
</tr>
<tr>
<td>Incorrect</td>
<td>50</td>
<td>50</td>
<td>50</td>
<td>150</td>
</tr>
<tr>
<td>Total</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>300</td>
</tr>
</tbody>
</table>

3.2. Annotation procedure

The 300 noun phrases were tagged by 4 annotators. The annotators were two experts (teachers of Spanish as a Foreign Language, who correct learners’ texts on a regular basis), which we will call E1 and E2, and two non-experts (native speakers of Spanish with higher education but without experience in corpus annotation), which we will call NE1 and NE2.

They all annotated the same noun phrase in the same sentences, but presented in different orders, using a Microsoft Excel spreadsheet. Annotators were provided with the target sentence plus the preceding and the following sentence, which they could resort to if they needed more context. They were asked to classify article usage for every noun phrase using one of the following tags: missing definite (AD), missing indefinite (AI), extraneous article (E), confusion error (C), difficult to judge (NC), article is correct (OK). They were not given any more guideline or training about the expected level of intervention in the texts: they were only asked to classify the noun phrases in one of the categories.

---

2 The texts were written by 4th grade Japanese students of Spanish with an intermediate level of proficiency, at Aichi Prefectural University.
4. Inter-annotator agreement

Tables 3 and 4 show the confusion matrices for expert and non-expert annotations. Observed agreement, defined as the proportion of items on which annotators agree, is 0.79 for expert annotators and 0.76 for non-experts.

<table>
<thead>
<tr>
<th>E1 ↓ E2→</th>
<th>AD</th>
<th>AI</th>
<th>C</th>
<th>E</th>
<th>NC</th>
<th>OK</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>AD</td>
<td>37</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>41</td>
</tr>
<tr>
<td>AI</td>
<td>0</td>
<td>5</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>7</td>
</tr>
<tr>
<td>C</td>
<td>0</td>
<td>0</td>
<td>30</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>36</td>
</tr>
<tr>
<td>E</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>39</td>
<td>7</td>
<td>1</td>
<td>50</td>
</tr>
<tr>
<td>NC</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>4</td>
<td>5</td>
<td>8</td>
<td>19</td>
</tr>
<tr>
<td>OK</td>
<td>4</td>
<td>0</td>
<td>4</td>
<td>7</td>
<td>10</td>
<td>122</td>
<td>147</td>
</tr>
<tr>
<td>Total</td>
<td>42</td>
<td>5</td>
<td>38</td>
<td>53</td>
<td>28</td>
<td>134</td>
<td>300</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>NE1 ↓ NE2→</th>
<th>AD</th>
<th>AI</th>
<th>C</th>
<th>E</th>
<th>NC</th>
<th>OK</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>AD</td>
<td>31</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>10</td>
<td>44</td>
</tr>
<tr>
<td>AI</td>
<td>2</td>
<td>5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>9</td>
</tr>
<tr>
<td>C</td>
<td>1</td>
<td>0</td>
<td>23</td>
<td>2</td>
<td>2</td>
<td>6</td>
<td>34</td>
</tr>
<tr>
<td>E</td>
<td>0</td>
<td>0</td>
<td>4</td>
<td>57</td>
<td>2</td>
<td>10</td>
<td>73</td>
</tr>
<tr>
<td>NC</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>OK</td>
<td>5</td>
<td>1</td>
<td>5</td>
<td>7</td>
<td>2</td>
<td>119</td>
<td>139</td>
</tr>
<tr>
<td>Total</td>
<td>39</td>
<td>8</td>
<td>32</td>
<td>68</td>
<td>6</td>
<td>147</td>
<td>300</td>
</tr>
</tbody>
</table>

However, using observed agreement to measure reliability does not take into account agreement that is due to chance and hence is not a good measure of reliability. Therefore, an analysis using Cohen’s Kappa statistic (Cohen, 1960) was performed. Perfect agreement would equate to a kappa of 1, and chance agreement would equate to 0. For the whole set of noun phrases (300, correct or incorrect), inter-annotator agreement for experts was found to be Kappa = 0.71 (p < 0.001), 95% CI (0.65, 0.77), and for non-experts it was 0.68 (p < 0.001), 95% CI (0.62, 0.75). If we exclude 45 sentences marked as “difficult to judge” by at least one annotator, kappa is 0.85 and 0.73 respectively. If we exclude 97 sentences tagged as correct by the four of them kappa is 0.62 and 0.58. If we exclude both sentences marked as NC by at least one annotator and sentences marked as OK by four annotators (remaining only 159 noun phrases all of them containing “safe” article errors) kappa is 0.79 and 0.61. Although kappa values vary depending on the set of sentences used to calculate it, agreement is over 0.60, which indicates “substantial agreement”.

These figures are slightly lower than those for English. In Han et al. (2006) annotators classify noun phrases in the same categories as our experiment with a kappa of 0.86, excluding correct noun phrases and sentences where they are unable to determine correct usage, which for us was 0.79 and 0.61 for experts and non-experts. The difference in the kappa values can in part be explained by the different proportion of article types in the data: while in our experiment article types are balanced (one third of noun phrases for every article type), in real texts like those used in Han et al. (2006) the zero category is the most common (followed by the definite and indefinite article) and this category also has the highest inter-annotator agreement3, which may raise the total kappa value.

In the following sections we examine different types of disagreement: disagreement due to the annotators’ individual biases (4.1), due to the annotation scheme (4.2) and genuine disagreement (4.3).

3 Full agreement (that is, by the four annotators) in sentences with an indefinite article is lower (45%) than in sentences with the zero article (71.0%) $\chi^2(4, N = 299) = 16.7$, p = 0.02.
4.1 Disagreement due to the annotators’ individual biases

As expected, non-expert annotators are less reliable than experts. First, non-expert annotators make more mistakes (they add tags which are incompatible with certain noun phrases, e.g. a missing article tag in a noun phrase already containing an article). To avoid this kind of mistakes, we should constrain the available tags depending on the input (e.g. if there is already an article in the noun phrase, do not allow the “missing” tag). Second, even though non-experts are supposed to be less confident on their annotation because pointing out errors in a text is a task for which they have no previous training, in fact they are less cautious than experts when they correct texts. This bias explains why, for example, NE1 uses the tag “difficult to judge” only one time (0.3%), while E2 uses it almost once every 10 sentences (9.3%), and non-experts use the tag “extraneous article” (specially for definite articles) more frequently than experts (23.5% vs 12.2% of times).

Part of the variability in annotators’ attitude could be reduced by giving clear guidelines about the optimum level of intervention in the texts. In this regard, we advocate for following a principle of minimal change: so we should not mark as errors the sentences where the learner choice is acceptable, even if the learner choice is not the best choice, that is, the goal of the annotator should be to produce an acceptable rather than a perfect result.

In relation to that, annotators should be instructed about the halo effect, by which the judgment of a sentence as acceptable or unacceptable is influenced by our overall impression of previous sentences. In other words, one is more likely to find errors in a text if this text already contains other errors. While expert annotators (teachers of a foreign language) are trained in evaluation methods and therefore they are aware of the importance of reliability in students’ evaluation, know how external factors (e.g. the halo effect and contrast effect) can have a negative impact and what can be done to reduce it, non-experts lack this training and do not know how to perform a fair evaluation -annotation. Therefore, non-expert annotators should receive training in evaluation methods to be able to reliably correct learner texts.

4.2 Disagreement due to the annotation scheme

With regard to the reliability of the 6 tags used for annotation, “difficult to judge” is the one that causes more disagreement: most of the times (67.7%) it is used by only one of the four annotators, and it is never used by three or four annotators in the same sentence. On the contrary, the rest of tags have a much higher agreement: on average, they are used by the four annotators 63.2% of the times, by three 19.9%, by two 9.2% and by one 7.7% of times. Therefore, this tag should at most be used to filter out problematic sentences, which annotators cannot comprehend, and not for proper annotation of sentences.

We advocate for not using this tag and instead set clear principles in the annotation guidelines specifying what the annotators should do when they are not confident about the error analysis of a sentence: exclude the sentence if a reasoned annotation is considered impossible (e.g. in incomprehensible fragments of text), or use more than one error tag if both are equally possible (as we will see in 4.3.2).

4.3 Genuine disagreement

Article presence and choice can be determined by different types of factors: it mainly depends on pragmatic factors (in our data, 69.0% of noun phrases), lexico-semantic factors (20.7%) and syntactic factors (10.3%).
As for pragmatic factors, for example the definite article is used to generalize, that is, to refer to a whole class of things or people, as in (1) (we underline the noun phrase and indicate the type of article in brackets) and to refer to something that is identifiable to the listener, as in (2). The indefinite is used to refer to any object of a particular class, as in (3), and no article is used when we are talking about an indefinite amount of something, as in (4) (examples from Alonso et al. (2013)).

(1) Los hijos dan muchos disgustos. [DEFINITE]
‘Children cause a great deal of trouble.’

(2) El hijo de María tiene dos años. [DEFINITE]
‘Maria’s son is two years old.’

(3) Tener un hijo es lo mejor que te puede pasar en esta vida. [INDEFINITE]
‘Having a child is the best thing that can happen in life.’

(4) No tengo hijos pero tengo sobrinos. [NO ARTICLE]
‘I do not have children but I have nephews.’

As for lexico-semantic factors, for example, place names usually have no article (México), while the definite is obligatory for rivers, mountains, seas and oceans (el Mediterráneo), and there exist many set phrases and idioms which require definite (e.g. con el objetivo de ‘with the objective of’), indefinite (por una parte, ‘on the one hand’) or zero article (e.g. a corto plazo, ‘in the short run’). As for syntactic factors, for example two or more nouns should have their own article if they refer to different things: un gato y un perro, “a cat and dog” (un gato y perro suggests a cross between a cat and a dog) (Butt and Benjamin, 2014).

Leaving aside sentences tagged as “correct” by 4 annotators, agreement is higher when the article choice depends on lexico-semantic factors (k = 0.835 for experts and 0.780 for non-experts) and lower with pragmatic factors ((k = 0.514 for experts and 0.496 for non-experts). Syntactic factors seem to be in between (k = 0.750 for experts and 0.523 for non-experts), although their low frequency makes the figures less reliable. Therefore, more care should be paid to pragmatic distinctions.

Specifically, disagreement is more likely in noun phrases where two pragmatic interpretations (and article choices) are possible, and annotators choose one of the alternatives in an inconsistent manner (§ 4.3.1 and § 4.3.2). Disagreement can also be due to a lack of the world knowledge that is needed to be able to determine the correct article usage (§ 4.3.3). As for syntactic and lexico-semantic factors (§ 4.3.4), disagreement occurs because annotators do not have a good knowledge about the existing prescriptive rules about article usage.

### 4.3.1. Vacillation between definite article and zero article

Frequently both the definite and zero article are acceptable for the same noun phrase. This happens when the noun phrase can refer to a whole class of things or people in general (definite article) as in (1) or to an indefinite amount of something (zero article) as in (4). This distinction frequently does not change the meaning of the sentence significantly and in fact some languages with articles like English usually use the zero article to express both situations.

In our experiment, when the two pragmatic interpretations are possible for a given sentence, annotators inconsistently choose one of them: some annotators tag the noun phrase for a missing

---

4 In (2) Maria’s son must be identifiable for the listener because a) Maria has only one son, or b) we have talked about him before.
article in (5) (OK|AD|AD|OK) while they tag it for an extraneous article in (6) (E |NC|OK|E), although both noun phrases can have the same pragmatic interpretations.

(5) Los políticos hablan en público y manifiestan sus opiniones con el objeto de conseguir votos de ciudadanos [...] [NO ARTICLE]

‘Politicians talk in public and show their opinion with a view to get votes from the citizens [...]’

(6) Concretamente los cursos que consiguieron participantes japoneses y que ofrecen los certificados oficiales como IMEC(Instituto de Medicina China) continuarán existiendo [...]. [DEFINITE]

‘Specifically the courses which obtained Japanese participants and offer official certificates like IMEC (Chinese Medicine Institute) will continue existing [...].’

In these cases, when both the definite and the zero article are acceptable, according to the principle of minimal change, we opt in favor of leaving the learners’ choice unchanged if it is acceptable.

4.3.2 Vacillation between indefinite article and zero article

Sometimes annotators agree in considering a noun phrase as unacceptable but they do not agree in the type of correction. This can happen when the learner wrongly uses a definite article, as in (8) (E|C|C|E), and the annotators propose different corrections: it can be an extraneous article if the noun phrase refers to an indefinite amount of something (zero article), or a confusion error if the noun phrase refers to any object of a particular class (indefinite).

(8) En cambio, la cocaína tiene el efecto tóxico. [DEFINITE]

‘On the contrary, cocaine has a toxic effect.’

When the two are equally acceptable and the annotator considers she cannot make a reasoned choice, we consider the best solution is to allow two error tags (E/C). In our experiment, this only happens with the pair of tags E and C.

4.3.3 Lack of world knowledge

In some sentences, annotators have insufficient extra-linguistic knowledge to be able to determine the right article usage. For example, in (9) (OK|E|E|E) the annotator needs to know whether in Nagoya there are only nine interesting and touristy places (definite article) or there are more than nine (no article).

(9) Sale cada treinta minutos aproximadamente desde la estación de Nagoya y paran en los nueve sitios muy interesantes y turísticos, por ejemplo El castillo de Nagoya. [DEFINITE]

‘It runs approximately every thirty minutes from Nagoya station and stops in nine very interesting and touristy places, for example Nagoya Castle.’

For future annotation, if the learner’s choice is acceptable in some context, as in (9), we do not mark it as wrong. If the learner’s choice is not acceptable, we tag the noun phrase as usual.

4.3.4 Lack of knowledge about syntactic and lexico-semantic rules

5 For example from the learner data, in parenthesis we indicate the tags chosen by the 4 annotators, in the following order: Expert 1, Expert 2, Non-expert 1, Non-expert 2. We also indicate in brackets the article choice of the learner.
Unlike article usage governed by pragmatic factors, which is subject to interpretation by the annotator, for article usage determined by syntactic and lexico-semantic constraints there exist some clear rules about what is considered correct and incorrect by the linguistic norm. These rules are part of language planning efforts by the Spanish language academy, but native speakers—even experts—do not have sufficient knowledge about them and as a result sometimes do not follow them when they annotate learner texts. For example, in (10) (AD|AD|OK|OK) experts marked as error an article usage that is actually accepted (RAE, 2006): the zero article between the preposition a (‘to’) and the relative pronoun que (‘which’).6

(10) [...] el capítulo 2 dice sobre el proceso del portuñol y los problemas a que el portuñol se enfrenta actualmente. [NO ARTICLE] ‘[...] chapter 2 is about the “portuñol” process and the problems that the “portuñol” confronts nowadays.’

Therefore, to determine the acceptability of article usage, annotators should not rely only on their intuition as native speakers but they should also consult existing rules and recommendations published in reference dictionaries and grammars as RAE (2006) and RAE (2009) to avoid contradictions between their corrections and what the linguistic norm actually says.

4. Corpus annotation

After analysing the main sources of disagreement, we have revised the annotation scheme as explained in Valverde & Ohtani (2014). Then, we have applied the revised annotation scheme to the annotation of an approximately 30,000 words sample of the CEDEL 2 learner corpus (Lozano & Mendikoetxea 2013), as shown in table 5. The texts in the sample were written without preparation, by learners whose first language is English.

Table 5. 30,000 words sample from the CEDEL2 learner corpus.

<table>
<thead>
<tr>
<th>Level</th>
<th>Words</th>
<th>Texts/Learners</th>
</tr>
</thead>
<tbody>
<tr>
<td>Beginner</td>
<td>10390</td>
<td>40</td>
</tr>
<tr>
<td>Intermediate</td>
<td>9960</td>
<td>22</td>
</tr>
<tr>
<td>Advanced</td>
<td>10293</td>
<td>20</td>
</tr>
<tr>
<td>Total</td>
<td>30643</td>
<td>82</td>
</tr>
</tbody>
</table>

The following categories were used: 1) Missing definite article, 2) Missing indefinite article, 3) Extraneous definite article, 4) Extraneous indefinite article, 5) Confusion error: indefinite instead of definite, 6) Confusion error: definite instead of indefinite, 7) Confusion error: another determiner instead of definite, 8) Confusion error: indefinite instead of another determiner.

Annotation has been carried out by one trained annotator with the software UAM Corpus Tool (O’Donnell 2010). We have found 196 errors in 30643 words, that is 0.64/100 words. Results are shown in Table 6. As expected, the most frequent error type involves the presence/absence of article: 92 missing articles (as in 10) and 95 extraneous articles (as in 11) give account of 95.41% of errors, and confusion errors represent only 4.59% (as in 12). This proportion is very close to that found in English learner texts: Han et al. (2006) found 21.5% of extraneous articles, 58.6% of missing articles and only 6.2% of a-the confusion in English texts written by Japanese learners. However, the frequency of extraneous articles in our texts—very close to missing articles—is higher than in the English texts.

6 The definite article is also acceptable but not obligatory. The definite article would be obligatory if the antecedent referred to a person, or if the subordinate clause was negative.
Table 6. Frequency of error tags by language level

<table>
<thead>
<tr>
<th>Error tag</th>
<th>Beginner</th>
<th>Intermediate</th>
<th>Advanced</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>%</td>
<td>N</td>
</tr>
<tr>
<td>Missing article</td>
<td>51</td>
<td>40.80</td>
<td>25</td>
</tr>
<tr>
<td>Extraneous article</td>
<td>68</td>
<td>54.40</td>
<td>12</td>
</tr>
<tr>
<td>Confusion error</td>
<td>6</td>
<td>4.80</td>
<td>2</td>
</tr>
<tr>
<td>Total</td>
<td>125</td>
<td>100</td>
<td>39</td>
</tr>
</tbody>
</table>

(10) Jude Law tiene pelo rubio y es Ingles. [0 → DEFINITE] ‘Jude Law has blond hair and is English’

(11) Fui en el junio y no llovó allí. [DEFINITE → 0] ‘I went in June and it did not rain there’

(12) Me encanta ir a la Universidad porque es la experiencia Buena. [DEFINITE → INDEFINITE] ‘I love going to the University because it is a good experience’

Among missing articles, the most frequent is the omission of the definite (88/92), as shown in Table 7. Among extraneous articles, the proportion of definite and indefinites is more balanced (37 vs 58).

Table 7. Frequency of error types by language level

<table>
<thead>
<tr>
<th>Error type</th>
<th>Beginner</th>
<th>Intermediate</th>
<th>Advanced</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>%</td>
<td>N</td>
</tr>
<tr>
<td>Missing type</td>
<td>51</td>
<td>100</td>
<td>25</td>
</tr>
<tr>
<td>Missing definite</td>
<td>50</td>
<td>98.04</td>
<td>25</td>
</tr>
<tr>
<td>Missing indefinite</td>
<td>1</td>
<td>1.96</td>
<td>0</td>
</tr>
<tr>
<td>Extraneous type</td>
<td>68</td>
<td>100</td>
<td>12</td>
</tr>
<tr>
<td>Extraneous indefinite</td>
<td>30</td>
<td>44.12</td>
<td>2</td>
</tr>
<tr>
<td>Extraneous definite</td>
<td>38</td>
<td>55.88</td>
<td>10</td>
</tr>
<tr>
<td>Confusion type</td>
<td>6</td>
<td>100</td>
<td>2</td>
</tr>
<tr>
<td>Definite instead of indefinite</td>
<td>5</td>
<td>83.33</td>
<td>0</td>
</tr>
<tr>
<td>Indefinite instead of definite</td>
<td>0</td>
<td>0.00</td>
<td>0</td>
</tr>
<tr>
<td>Indef. instead of another det.</td>
<td>0</td>
<td>0.00</td>
<td>1</td>
</tr>
<tr>
<td>Another det. instead of definite</td>
<td>1</td>
<td>16.67</td>
<td>1</td>
</tr>
</tbody>
</table>

From this data we can extract some statistically significant differences among learners. For example, for beginners learners extraneous articles (54.40%) are more frequent than missing articles (40.80%), while for intermediate learners the opposite is true: missing articles (64.10%) are more frequent than extraneous ones (30.77%). As for the type of missing article, among beginner learners the indefinite is very rare (1.96%), but not so rare for advanced learners (18.75%).

5. Conclusions

Although article errors have been annotated in a number of small-scale studies, to date there has not been any study about article error annotation and inter-annotator agreement in Spanish learner texts. In this paper we have tested the results of an annotation scheme for article errors in a sample of learner texts written by Japanese learners. We have calculated agreement among 4 annotators (2 experts and 2 non-experts) and have found kappa values between 0.85 and 0.62 for expert annotators and from 0.73 to 0.58 for non-experts, depending on the collection of sentences considered.
Non-experts are less reliable than experts, and the annotation scheme (the tag “difficult to judge”) is also responsible for part of the disagreement.

As for genuine disagreement among annotators, some pragmatic distinctions are specially problematic: the distinction between a) a whole class of things or people in general (definite article) and b) an indefinite amount of something (zero article), and the distinction between a) an indefinite amount of something (zero article) and any object of a particular class (indefinite article). In addition to that, some times more world knowledge is needed to determine whether article presence and choice is acceptable or not. As for article usage governed by syntactic and lexico-semantic factors, annotators sometimes disagree in determining the right article usage because they lack knowledge about the existing prescriptive rules published by the Spanish language academy.

To improve annotation reliability, annotators need to be trained in language evaluation methods and have to consult published prescriptive rules about article usage. After annotating a 30,000 words sample from the CEDEL2 learner corpus with a revised annotation scheme, we have found that the most frequent error types are missing and extraneous articles.

Acknowledgements

This work was supported by kakenhi (25770207), Grant-in-Aid for Scientific Research from the Japan Society for the Promotion of Science.

References


Tools for Supporting Language Acquisition via Extensive Reading

Alexandra UITDENBOGERD*
*School of Computer Science and IT, RMIT University, Australia
*alexandra.uitdenbogerd@rmit.edu.au

Abstract: Extensive reading, that is, the reading of large quantities of text at a comfortable level of difficulty, has been shown to be of great benefit to second and foreign language learners' skills. We define seven types of reading support as management, text generation, text selection, text simplification, preparation, translation, and revision. We describe and propose a range of tools that assist learners to locate reading material of an appropriate level of difficulty and manage their progress. Our prototype implementation of two of these tools, the Bilingual eReader and the Readable Extract Search Engine, demonstrate their feasibility.

Keywords: CALL, extensive reading

1. Introduction

Imagine attempting to make sense of the following extract while reading:

She was the youngest of the two daughters of a most XXX, XXX father; and had, in XXX of her sister's XXX, been XXX of his house from a very early XXX.

The “XXX”s above represent unknown words. For this extract, you are reading a text in which you know 81.25% of the words. If you are a native or highly fluent speaker of English, you have the advantage of excellent knowledge of the grammatical flow of English sentences, which will greatly assist you in your ability to predict the meaning of the missing words. Research has shown that for most foreign language learners to predict the meaning of words in text requires knowing at least 95% of the words (Liu and Nation, 1985). However, for most native texts, the learner would need a vocabulary of 5,000 word families in order to achieve that coverage (Hirsh and Nation, 1992). A typical 1,000 hour course of English as a second language will achieve a 2,000 word vocabulary (Laufer, 2000). This leaves a substantial shortfall in vocabulary knowledge.

It is very challenging to become proficient in a foreign language to the extent that it can be used for discourse at a high level. One useful activity for improving language skills that is readily available is extensive reading. However, the reading material needs to be at an appropriate level of difficulty as well as interesting to be the most effective. For English and French there is an extensive collection of reading material available, such as the Oxford Bookworms series. Other languages are less resource rich. However, there are large quantities of text produced in a variety of languages on the Web and elsewhere. A proportion of these are potentially suitable for reading practice. Uitdenbogerd (2006) examined a corpus of English web sites and found that the readability range of Web-sites adequately covered that of typical stories written for English language learners. However, many such web-sites provided uninteresting reading, such as navigation pages. Heilman et al. (2010) reached a similar conclusion in their work, and developed methods to eliminate web pages that don’t have sufficient prose, as well as providing categories for students to choose from to increase the chance that the material retrieved would be interesting to them.
The contributions of this paper are:

1. a taxonomy of extensive reading support that can be provided by computer (and other) systems (See Section 3)
2. tools and techniques for providing appropriate text from a corpus of literature (See Section 4)

2. Background

The need for language skills is becoming more relevant in globally connected societies. For example, millions of Chinese students are studying English (Zheng and Cheng, 2008), and 27 million people have taken the TOEFL English language test (TOEFL website). It is generally agreed that more exposure to language improves skills in the language. For example, various studies have shown vocabulary gain from reading (Waring and Takaki, 2003).

In a previous paper we observed that there are short extracts and sentences to be found in the classical French literature that meet strict vocabulary criteria such as consisting only of the 20 most frequently occurring words in news text, French-English cognates, and proper nouns (Uitdenbogerd, 2010). We provided an estimate of exact cognates in native text (10%) and a frequency distribution of sentence structures.

Clearly the more strict the constraints on the text, the fewer suitable extracts will be found. However, even at the strictest constraints such as 1-word sentences, or sentences consisting only of the most frequent 20 words, proper nouns and French-English cognates, extracts could be found. Relaxing constraints to allow 95% coverage provides larger quantities of extracts.

3. Extensive Reading Support Taxonomy

Support for extensive reading falls into seven main categories: management of an extensive reading-based programme, generation of readable text, selection of readable text, simplification of text, learner preparation, glossary or translation support, and revision.

- **Management** via an extensive reading system frees the user from tracking their reading and their progress. It can work in conjunction with text selection and other categories of support by overseeing the different types of activity available to the user.

- **Text Generation** involves generating stories or other content based on strict readability constraints. For early stages that require much reading practice with a small set of vocabulary and grammar this can be a useful addition to the choices available for reading.

- **Text Selection** via readability-based search enables the learner to find easier texts to read from large corpora. When combined with topic search, texts can be both relevant and readable. Text selection can also be based on recommendations in the manner of typical ratings-based recommender systems. Text selection support can help ensure that text is both readable and interesting, increasing the motivation of the learner to read and benefit from reading.

- **Text Simplification** has traditionally been done manually, typically for classic works of literature to make them accessible to both children and foreign language learners. Typically the works are not only simplified in language, but made significantly shorter.

- **Preparation** consists of materials that assist the learner with vocabulary and background knowledge before they commence reading. (See Section 4.1)

- **Translation** of difficult vocabulary occurring in text helps the learner to read fluently, and when a gloss is used to look up a word, the word is more remembered than if it is merely read over (Lomicka, 1998). Translations of difficult passages can also help a learner tackle more difficult texts.

- **Revision** typically consists of a set of questions that the learner answers once they have completed reading a text. The questions test the learner’s comprehension of the text, as well as their knowledge of vocabulary and grammar seen in the text. The more involved the learner is with the language, the more they will remember. Therefore, exercises based on the reading will improve language retention (Laufer 2000).
Applications vary in their support across these categories. In Section 5 we present a range of applications and discuss them in relation to the categories of support and their applicability to language acquisition via extensive reading.

4. Applications

We discuss various existing and proposed applications that provide support of the types described in our taxonomy.

4.1 The Readable Document Search Engine

An idea that was independently developed by several research groups, and is now a part of major search engines is to make readability a criterion for search for documents. The most developed and used system that retrieves documents for foreign language reading practice is the REAP system (Heilman et al. 2010). It allows instructors to choose target vocabulary to be studied by students, permits students to select broad interest areas, and presents the students with recommended texts of an appropriate reading level that provide practice in the target vocabulary. REAP determines the student’s existing vocabulary, so that it can better estimate the reading level required. While it doesn’t appear to provide preparation activities, it does provide comprehensive translation support by making any word searchable in an on-line dictionary, and target words for study are hyperlinked to a definition. Revision consists of a practice exercises on target words. The system has been shown to improve language knowledge.

Where systems recommend native texts and also track a student’s vocabulary knowledge, a personalised set of preparation material could be provided. For example, based on the word frequencies in the document and the known vocabulary, a set of 5-10 words that would improve the student’s ability to understand the document the most, could be presented in a small pre-reading lesson. A simple but not necessarily optimal approach would be to select the highest frequency document words that are not in the student’s current known vocabulary. These are likely to be “topic” words. For example, in a story about pirates, the words “pirate”, “sail”, “mast”, “anchor” and “cabin” may occur frequently, but be generally unknown by a student studying the English Language. Where personalised vocabulary tracking doesn’t exist, the preparation material can be selected purely by the frequency of words in the document versus their frequency in background text. Some published reading books for foreign language learners do this to some extent.

4.2 The Readable Extract Search Engine

The idea behind the readable extract search engine is that a collection of native texts that are of high quality and interest, such as a collection of literature, may provide short extracts for reading. In an initial exploration we found it was possible to locate short readable extracts based on both vocabulary and grammar constraints (Uitdenbogerd, 2010). In our prototype we used sentence length as the readability measure, which has been shown to work very well for the case of French as a foreign language for English speakers (Uitdenbogerd, 2006). However, a more sophisticated readability measure, or one that is appropriate for a different language can be substituted.

Users set the sentence length criteria and the ideal extract size (Figure 1), and then retrieve a list of results (Figure 2). Users can then choose which extract is of interest from a list of titles and extract lengths.

In our prototype implementation, the extract is shown highlighted within the full text, allowing users to continue reading if they so choose (Figure 3). While the usefulness of short extracts hasn’t been demonstrated as yet, extracts that provide multiple examples of a particular word or phrase has increased vocabulary knowledge (Webb, 2007).
Figure 1. The Readable Extract Search Engine splash screen.

Figure 2. Retrieved list of extracts.
4.3 The Bilingual e-Reader

The bilingual e-Reader combines simplification with translation, in that the more difficult sentences are presented in the learner’s native language. The idea of mixed language reading material is not new. The approach has been used for gradually increasing the amount of target language in a story for English-German with good results compared to a normal German lesson (Weible, 1980), and has also been used to introduce Japanese kanji within an English story (Watanabe, 2002).

We developed a simple prototype using sentence length as the readability criterion, with a collection of movie subtitles (downloaded from http://opus.lingfil.uu.se/OpenSubtitles.php) as the corpus. Figure 4 shows the main screen, including parameters for source and target language, a list of movie subtitles available, a readability parameter (maximum sentence length), and the resulting mixed language subtitles for the movie Rocky with the maximum sentence length set to 7.

The idea is a little controversial in the sense that it prevents total immersion in the target language, but it can be useful for earlier stages of language learning. Later stages would benefit from sentences that translate when clicked on. This way translations only are shown when requested.
4.4 The Book Bootstrapper

This application re-orders the text of a book into readability order. This is a similar idea to the Textladder application that sorts a collection of documents into readability order (Ghadirian, 2002), but is done with a single large text. Fiction books with significant amounts of dialogue have many simple sentences and considerable variation in readability across the text. For example, the novel *Emma* by Jane Austen has 31 occurrences of the sentence “Oh!”, 6 of the sentence “Ah!”, and many sentences consisting of a single-word name. Its average sentence length is approximately 21.

If it is true that “meeting” a word 10 times during reading allows one to learn its meaning, then reading a novel of about 85,000 words should provide a learner with a 1,000 word reading vocabulary (Waring, 2009). However, the learner must be able to comprehend enough of the text to gain vocabulary from it. Ordering the text on readability criteria may allow vocabulary to be acquired more comfortably than reading it in the normal order. The learner can then re-read the book in the normal order with greater ease. We have developed a prototype book bootstrapper, and are currently determining its range of usefulness.

4.5 The Text Simplifier

An alternative to the careful selection of text, or the efforts of writing simple text, is the automatic text simplifier. Using similar techniques to document summarisation in addition to word substitution, the text simplifier both shortens a given text and reduces the vocabulary and grammar difficulty. The usefulness of simplification over translation was demonstrated by Eskenazi, Lin and Saz (2013).

4.6 The Text Generator

For extensive reading practice, the text generator should generate interesting stories on a small vocabulary and grammar repertoire. Shim and Kim (2002) developed a story generator that uses autonomous agents. To our knowledge, this idea hasn’t been applied to foreign language learning yet. While not exclusively reading-based, chat-bots can also provide engaging reading practice.
5. Conclusion and Future Work

Reading extensively in the target language improves language skills, and this happens most efficiently when the reading material is both interesting and at a suitable difficulty level. Therefore systems that can provide appropriate reading material in sufficient quantities will be of great benefit to the language learner. We identified seven areas of support that applications can provide for extensive reading: management, generation of readable text, selection of readable text, simplification of text, learner preparation, glossary or translation support, and revision.

We described simple prototype systems that allow the user to locate suitable reading material, either as extracts, or as mixed language texts, exploiting the variability in language difficulty across a typical native text. Other systems were described that are either already in existence or may be worthwhile additions to the range of tools for the language learner. We are currently determining the applicability of one of these: the book bootstrapper. Future work will include determining the effectiveness of the proposed applications for language learning.

Acknowledgements

Software development for the: Readable Extract prototype was by Christopher Dore, Len Wang and XinYang Yao; and the Bilingual eReader was by Wei Yan, Yueyao Zhuang and Ziteng Zhang. The Book Bootstrapper is currently being developed by Daniel Collins.

The Bilingual eReader concept was originally discussed with Laurianne Sitbon, who suggested the use of movie subtitles.

Support for this project has been provided by the Australian Government Office for Learning and Teaching. The views in this project do not necessarily reflect the views of the Australian Government Office for Learning and Teaching.

References


Waring, R. (2009). The inescapable case for extensive reading. Extensive Reading in English Language Teaching, Andrzej Cirocki (ed.).


Overview of Grammatical Error Diagnosis for Learning Chinese as a Foreign Language

Liang-Chih YU\textsuperscript{a,b}, Lung-Hao LEE\textsuperscript{c,d}* & Li-Ping CHANG\textsuperscript{e}

\textsuperscript{a}Department of Information Management, Yuen-Ze University, Taiwan
\textsuperscript{b}Innovation Center for Big Data and Digital Convergence, Yuen-Ze University, Taiwan
\textsuperscript{c}Information Technology Center, National Taiwan Normal University, Taiwan
\textsuperscript{d}Department of Computer Science and Information Engineering, National Taiwan University, Taiwan
\textsuperscript{e}Mandarin Training Center, National Taiwan Normal University, Taiwan

\*lhlee@nlg.csie.ntu.edu.tw

Abstract: We organize a shared task on grammatical error diagnosis for learning Chinese as a Foreign Language (CFL) in the ICCE-2014 workshop on Natural Language Processing Techniques for Educational Applications (NLPTEA). In this paper, we describe all aspects of this shared task, including task description, data preparation, evaluation metrics, and testing results. The aim is, through such evaluation campaigns, more advanced computer-assisted Chinese learning techniques will be emerged.

Keywords: Computer-assisted language learning, shared task, Mandarin Chinese

1. Introduction

China’s growing global influence has prompted a surge of interest in learning Chinese as a foreign language (CFL), and this trend is expected to continue. However, whereas many computer-assisted learning tools have been developed for use by students of English as a Foreign Language (EFL), support for CFL learners is relatively sparse, especially in terms of tools designed to automatically detect and correct Chinese grammatical errors. For example, while Microsoft Word has integrated robust English spelling and grammar checking functions for years, such tools for Chinese are still quite primitive.

In contrast to the plethora of research related to EFL learning, relatively few studies have focused on computer-assisted language learning for CFL learners. Relative position and parse template language models have been adopted to detect Chinese errors written by US learners (Wu et al. 2010). Machine learning models have been applied to detect word-ordering errors in Chinese sentences from the HSK dynamic composition corpus (Yu and Chen, 2012). Ranking SVM based model has been further explored to rank the candidates and suggest the proper corrections of word ordering errors (Cheng et al. 2014). A penalized probabilistic First-Order Inductive Learning (pFOIL) algorithm has been proposed for grammatical error diagnosis (Chang et al. 2012). Linguistic rule based approach has been presented to detect grammatical errors written by CFL learners (Lee et al. 2013). A sentence judgment system has been implemented to integrate rule-based linguistic analysis and n-gram statistical learning for detecting grammatical errors (Lee et al. 2014). SIGHAN 2013 bakeoff on Chinese spelling check evaluation focus on developing automatic checker to detect and correct spelling errors (Wu et al. 2013). In summary, human language technologies for Chinese learning have attracted more attentions in recent years.

In the ICCE-2014 workshop on Natural Language Processing Techniques for Educational Applications (NLPTEA), we organize a shared task on Chinese grammatical error diagnosis that provides an evaluation platform for developing and implementing computer-assisted learning tools. The data sets in our task are collected from the Chinese as the Foreign Language (CFL) learners’ written essays. Given a sentence with/without one of grammatical errors, \textit{i.e.}, redundant word, missing word, word disorder, and word selection, the developed system should indicate whether contains grammatical errors and further points out which one of defined error types. The hope is that,
through such evaluation campaigns, more advanced Chinese grammatical error detecting techniques will be emerged.

We give an overview of the shared task on grammatical error diagnosis for learning Chinese as a foreign language. The rest of this article is organized as follows. Section 2 details the designed task. Section 3 introduces the data sets provided in this evaluation. Section 4 proposes the evaluation metrics. Section 5 presents the results of participants’ approaches for performance comparison. Finally, we conclude this paper with the findings and future research direction in the Section 6.

2. Shared Task Description

The goal of this shared task is developing the computer-assisted tools to detect several kinds of grammatical errors, that is, redundant word, missing word, word disorder, and word selection. The input sentence contains one of defined error types. The developed tool should indicate which kind of error type is embedded in the given sentence. If the input sentence, which is given a unique sentence number SID, contains no grammatical errors, the tools should return “SID, Correct”. If an input sentence contains a defined grammatical error, the output format should be “SID, error_type”. We simplify the task that there are only one error type may be in the given sentence. Examples are shown as follows. In example 1, the character “被” is a redundant word. There is a missing word “有” in the example 2 and its correct usage is shown in example 3. The sentence in the example 4 has word disorder error, i.e., the word “很早” should be preceded the word “起床”. The word “一個” in the example 5 is an incorrect word selection, the correct word should be “一件”.

- **Example 1**
  Input: (sid=B2-1447-6) 希望沒有人再被食物中毒
  Output: B2-1447-6, Redundant
- **Example 2**
  Input: (sid=C1-1876-2) 對社會國家不同的影響
  Output: C1-1876-2, Missing
- **Example 3**
  Input: (sid=C1-1876-2) 對社會國家有不同的影響
  Output: C1-1876-2, Correct
- **Example 4**
  Input: (sid=A2-0775-2) 我起床很早
  Output: A2-0775-2, Disorder
- **Example 5**
  Input: (sid=B1-0110-2) 我會穿著一個黃色的襯衫
  Output: B1-0110-2, Selection

3. Data Sets

Mandarin Training Center (MTC) of National Taiwan Normal University (NTNU) was founded in 1956 for teaching Chinese as a foreign language. Currently, MTC is the most renowned Chinese language center in Taiwan, in which around 1700 CFL learners from more than 70 countries enrolled each academic quarter. The learner corpus used in our task is collected from the computer-based writing Test of Chinese as a Foreign Language (TOCFL). The writing test is designed according to the six proficiency levels of the Common European Framework of Reference (CEFR). Test takers have to complete two different tasks for each level. For example, for the A2 (Waystage level) candidates, they will be asked to write a note and describe a story after looking at four pictures. All candidates are asked to complete the writings on line.

We further ask the annotators to label the grammatical errors in CFL learners’ written sentences and provide their correct usage. Our prepared data is further divided into three distinct sets.

1. **Training set**: 1,506 CFLs’ writings are collected in which 5,607 grammatical errors are annotated. Each CFL learners’ writing is represented in SGML format shown in Figure 1. The title attribute is
used to describe the topic of the writing test. There is only one grammatical error in an annotated sentence. The error types are also indicated along with their corresponding correct usages. All sentences in this set can be used to train the developed grammatical error detection tool. (2) **Dryrun set**: Total 33 sentences are given for participants to familiarize themselves with the final testing process. Each participant can submit several runs generated using different models with different parameter settings. In addition to make sure the submitted results can be correctly evaluated, participants can fine-tune their developed models in the dryrun phase. The purpose of dryrun is for output format validation only. No matter which performance can be achieved that will not be included in our official evaluation. (3) **Test set**: In total, there are 1,750 testing sentences. A half of these instances contain no grammatical errors. Another half of testing cases includes one grammatical error per sentence. The number of error type redundant, missing, disorder, and selection is 279, 350, 120, and 126, respectively. The distribution is the same with our given training set. The policy of our evaluation is an open test. In addition to our provided data sets, registered research teams can employ any linguistic and computational resources to detect grammatical errors in the sentences.

![Figure 1. An essay represented in SGML format.](image)

### 4. Performance Metrics

Table 1 shows the confusion matrix used for performance evaluation. In the matrix, True Positive (TP) is the number of sentences with grammatical errors that are correctly proposed by the developed tool; False Positive (FP) is the number of sentences without grammatical errors that are incorrectly proposed; True Negative (TN) is the number of sentences without grammatical errors that are identified correctly; False Negative (FN) is the number of sentences with grammatical errors that are incorrectly regarded as correct sentences.

<table>
<thead>
<tr>
<th>Confusion Matrix</th>
<th>System Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>Positive (With grammatical errors)</td>
<td>Negative (Without grammatical errors)</td>
</tr>
<tr>
<td>Gold Standard</td>
<td></td>
</tr>
<tr>
<td>Positive</td>
<td>True Positive (TP)</td>
</tr>
<tr>
<td>Negative</td>
<td>False Positive (FP)</td>
</tr>
</tbody>
</table>

The criteria for judging correctness are distinguished into two levels. (1) **Detection level**: all error types are regarded as incorrect. Binary classification of a testing instance, *i.e.*, correct or incorrect, should be completely identical with the gold standard. (2) **Identification level**: this level could be considered as a multi-class categorization problem. In addition to correct instances, all error
types should be clearly identified, i.e., Redundant, Missing, Disorder, and Selection. The following metrics are measured in both levels with the help of the confusion matrix.

- **False Positive Rate (FPR)** = \( \frac{FP}{FP+TN} \)
- **Accuracy** = \( \frac{TP+TN}{TP+FP+TN+FN} \)
- **Precision** = \( \frac{TP}{TP+FP} \)
- **Recall** = \( \frac{TP}{TP+FN} \)
- **F1** = \( 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \)


- **False Positive Rate (FPR)** = 0.5 (= 2 / 4).

- **Detection Accuracy** = 0.75 (=6/8).

- **Detection Precision** = 0.67 (=4/6).

- **Detection Recall** = 1 (=4/4).

- **Detection F1** = 0.8024 (=2*0.67*1/(0.67+1))

- **Identification Accuracy** = 0.5 (=4/8).

- **Identification Precision** = 0.33 (=2/6).

- **Identification Recall** = 0.5 (=2/4).

- **Identification F1** = 0.3976 (=2*0.33*0.5/(0.33+0.5))

5. Evaluation Results

Table 2 shows the participant teams and their testing submission statistics. Our shared task attracted 13 research teams. There are 4 teams that come from Taiwan, i.e., AS, KUAS & NTNU, NCYU, and NTOU. 3 teams originate from China, i.e., HITSZ, PKU, and PolyU. The remaining 6 teams are CIRU from United States of America, MU from New Zealand, SPBU from Russia, TMU from Japan, UL from United Kingdom, and UDS from Germany. Among 13 registered teams, 6 teams submitted their testing results. In total, we had received 13 runs in the formal testing phase.
Table 2: Result submission statistics of all participants.

<table>
<thead>
<tr>
<th>Participants (Ordered by abbreviations of names)</th>
<th>#Submissions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Academia Sinica (AS)</td>
<td>0</td>
</tr>
<tr>
<td>Confucius Institute of Rutgers University (CIRU)</td>
<td>1</td>
</tr>
<tr>
<td>Harbin Institute of Technology Shenzhen Graduate School (HITSZ)</td>
<td>0</td>
</tr>
<tr>
<td>National Kaohsiung University of Applied Sciences &amp; National Taiwan Normal University (KUAS &amp; NTNU)</td>
<td>3</td>
</tr>
<tr>
<td>Massey University (MU)</td>
<td>0</td>
</tr>
<tr>
<td>National Chiayi University (NCYU)</td>
<td>1</td>
</tr>
<tr>
<td>National Taiwan Ocean University (NTOU)</td>
<td>2</td>
</tr>
<tr>
<td>Peking University (PKU)</td>
<td>0</td>
</tr>
<tr>
<td>The Hong Kong Polytechnic University (PolyU)</td>
<td>0</td>
</tr>
<tr>
<td>Saint Petersburg State University (SPBU)</td>
<td>0</td>
</tr>
<tr>
<td>Tokyo Metropolitan University (TMU)</td>
<td>2</td>
</tr>
<tr>
<td>Saarland University (UDS)</td>
<td>4</td>
</tr>
<tr>
<td>University of Leeds (UL)</td>
<td>0</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>13</strong></td>
</tr>
</tbody>
</table>

Table 3: Testing results of our shared task.

<table>
<thead>
<tr>
<th>Submission</th>
<th>FPR</th>
<th>Detection Level</th>
<th>Identification Level</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Acc.</td>
<td>Pre.</td>
</tr>
<tr>
<td>CIRU-Run1</td>
<td>0.496</td>
<td>0.6446</td>
<td>0.6128</td>
</tr>
<tr>
<td>KUAS &amp; NTNU-Run1</td>
<td>0.904</td>
<td>0.5006</td>
<td>0.5003</td>
</tr>
<tr>
<td>KUAS &amp; NTNU-Run2</td>
<td>0.2686</td>
<td>0.5217</td>
<td>0.5374</td>
</tr>
<tr>
<td>KUAS &amp; NTNU-Run3</td>
<td>0.904</td>
<td>0.5006</td>
<td>0.5003</td>
</tr>
<tr>
<td>NCYU-Run1</td>
<td><strong>0.1189</strong></td>
<td>0.4983</td>
<td>0.4927</td>
</tr>
<tr>
<td>NTOU-Run1</td>
<td>1</td>
<td>0.5</td>
<td>0.5</td>
</tr>
<tr>
<td>NTOU-Run2</td>
<td>1</td>
<td>0.5</td>
<td>0.5</td>
</tr>
<tr>
<td>TMU-Run1</td>
<td>0.1977</td>
<td>0.5171</td>
<td>0.5399</td>
</tr>
<tr>
<td>TMU-Run2</td>
<td>0.1691</td>
<td>0.5103</td>
<td>0.5287</td>
</tr>
<tr>
<td>UDS-Run1</td>
<td>0.792</td>
<td>0.4914</td>
<td>0.4945</td>
</tr>
<tr>
<td>UDS-Run2</td>
<td>0.6286</td>
<td>0.4949</td>
<td>0.4959</td>
</tr>
<tr>
<td>UDS-Run3</td>
<td>0.5783</td>
<td>0.4949</td>
<td>0.4955</td>
</tr>
<tr>
<td>UDS-Run4</td>
<td>0.2491</td>
<td>0.5046</td>
<td>0.509</td>
</tr>
</tbody>
</table>

Table 3 shows the testing results of our shared task. In addition to achieving promising detection effects of grammatical errors, reducing the false positive rate, which is percentage of the correct sentences that are incorrectly reported containing grammatical errors, is also important. The research teams, NCYU and TMU, achieved relatively low false positive rates.

Detection level evaluations are designed to detect whether a sentence contains grammatical errors or not. Accuracy is usually adopted to evaluate the performance, but it is affected by the distribution of testing instance. The baseline can be achieved easily by always guessing without errors. That is accuracy of 0.5 in this evaluation. Some systems achieved slightly better than the baseline, i.e., CIRU, KUAS&NTNU, TMU and UDS. Registered teams may send different runs that aimed at optimizing the recall or precision rates. These phenomena guide us to adopt F1 score to
reflect the tradeoff between precision and recall. In the testing results, CIRU accomplished the best detection effects of indicating grammatical errors, which resulted the best F1 score 0.6884. For identification level evaluations, the systems need to identify the error types in the given sentences. The research team came from CIRU accomplished the best correction accuracy 0.4589. Most systems cannot effectively identify the input sentences to point out possible grammatical errors. Our testing results indicate that the system developed by CIRU accomplished the best identification F1 0.4333.

In summary, it is a really difficult task to develop the computer-assisted learning tool for grammatical error diagnosis, especially learning Chinese as a foreign language, since there are only target sentences without the help of their context. We cannot find a relatively promising system according to our testing results. In general, this research problem still has long way to go.

6. Conclusions and Future Work

This paper describes the overview of NLPTEA 2014 shared task on grammatical error diagnosis for learning Chinese as a foreign language. We introduce the task designing ideas, data preparation details, evaluation metrics, and the results of performance evaluation. This task also encourages researchers to bravely propose various ideas and implementations for possible breakthrough. No matter how well their implementations would perform, they contribute to the community by enriching the experience that some ideas or approaches are promising or impractical, as verified in this shared task. Their reports in the proceeding will reveal the details of these various approaches and contribute to our knowledge about computer-assisted Chinese learning.

All data sets and their accompanying gold standards and evaluation tool are publicly available for research purposes at http://ir.itc.ntnu.edu.tw/lre/nlpteal4cfl.htm. We hope our provided data can serve as a benchmark to help developing better Chinese learning tools. This shared task also motivates us to build more language resources in the future to possibly improve the state-of-the-art techniques.

Acknowledgements

This research was supported by the Ministry of Science and Technology, under the grant MOST 102-2221-E-155-029-MY3, 103-2221-E-003-013-MY3, 103-2911-I-003-301 and the “Aim for the Top University Project” and “Center of Learning Technology for Chinese” of National Taiwan Normal University, sponsored by the Ministry of Education, Taiwan.

References


KNGED: a Tool for Grammatical Error Diagnosis of Chinese Sentences

Tao-Hsing CHANG*, Yao-Ting SUNGb, Jia-Fei HONGc & Jen-I CHANGb

*Department of Computer Science and Information Engineering, National Kaohsiung University of Applied Sciences, Taiwan
bDepartment of Educational Psychology and Counseling, National Taiwan Normal University, Taiwan
cDepartment of Applied Chinese Language and Culture, National Taiwan Normal University, Taiwan
*changth@gm.kuas.edu.tw

Abstract: The main purpose of this paper is to propose a method that can automatically detect whether there are any grammatical errors as well as identify their error types. The framework of this method is based on a rule base to identify common grammatical errors. This rule base contains manually constructed rules and rules that are automatically machine generated. This paper further proposes algorithms which can apply these rules to determine whether a sentence is incorrect as well as what types of errors it belongs to. Experimental results show that the F1-measure of the proposed method is 0.64 and 0.30 on detection and identification, respectively.

Keywords: KNGED, Chinese grammar, grammatical error, CFL, automatic diagnosis, rule-based method.

1. Introduction

Automatically detecting grammatically incorrect sentences is fundamental and important for numerous NLP studies and related applications. For instance, in language teaching, automatic detection of grammatically incorrect sentences produced by learners can help a teacher teach grammar more effectively. However, the detection of grammatically incorrect sentences in Chinese is challenging. The main reason is that it is difficult to detect sentence boundaries in Chinese. In English, the period provides a clear signal of the end of a sentence, allowing the segment of text between two periods to be taken as a sentence and analyzed grammatically. However, two periods between Chinese sentences represent a complete semantic expression. An excessively long sentence may contain several commas as delimiters. Moreover, a sentence segment formed by a comma may comprise the complete sentence, a clause and even a phrase. This phenomenon makes the detection of sentence boundaries in Chinese difficult.

The above difficulty makes the method of detecting grammatical errors in Chinese sentences via using a parser to completely deconstruct a parsing tree infeasible. To detect errors in an English sentence, a parsing tree constructed using a parser can provide criteria for sentence judgment. However, to deal with grammatical errors generated by learners of Chinese as a second language, the parsing tree method may not have been thoroughly examined. This is because the main cause of common errors by second language learners is the language transfer phenomenon in language learning. For example, Korean students often write the following incorrect sentence.

我 來 台 灣 四年 工 作 了

The error pattern here is that the time noun ‘四年’ appears before the verb ‘工作’. This common error occurs because Korean is a subject-object-verb (SOV) language. An important characteristic of an SOV language is that all other elements such as nouns, adverbs, and numbers come before the verb. However, Chinese is a subject-verb-object (SVO) language, so the correct
sentence would be ‘我來台灣四年工作了 (I came to work in Taiwan for four years)’, where the verb ‘工作(work)’ must appear before the time noun ‘四年 (four years)’.

Consequently, by generalizing common errors made by learners of Chinese as a second language, it is possible to further analyze which specific syntactic structures those common errors belong to. These specific syntactic structures are considered error detection rules. Moreover, the error detection rules possess a syntactic error pattern and its corresponding syntactically correct pattern. Sufficient patterns and rules enable the generation of a rule base. Sentences can be compared using the rule base to identify grammatical errors. If a sentence contains multiple segments that conform to error detection rules, this segment is most likely the syntactic structure of that error. Therefore, sufficient error detection rules are collected, grammatical errors can be identified by comparing the rule base. The method of error detection rule collection can be obtained through analysis of sentences in a learner error corpus. The bigger the corpus, the more error detection rules can be generated and the more grammatical errors detected.

This study primarily proposes a method capable of automatically detecting and identifying grammatical errors. The framework of this method is based on a rule base to identify common grammatical errors. This rule base contains manually constructed rules and rules that are automatically machine generated. This study further proposes algorithms that can apply error detection rules to determine whether a sentence is incorrect and what types of errors it belongs to and classify any errors.

This paper is organized as follows. Section 2 reviews some related research and illustrates the impact of these studies on the research motivations. Section 3 then lists the corpus used in this paper and illustrates a learner corpus that is designed to automatically detect grammatically incorrect sentences produced by Chinese as second language writers. Next, section 4 introduces the method to manually construct error detection rules and the method for the program to automatically generate error detection rules. Subsequently, section 5 describes the algorithm for automatically identifying incorrect sentences. Section 6 demonstrates the performance of the proposed approach, while Section 7 draws conclusions.

2. Related Works

Recently, Chinese learning has become a growing trend, making Chinese one of the most popular foreign languages globally, besides English. To learners, learning a new language frequently involves grammar difficulties, and grammatically incorrect sentences are a common error. Previous research on second language acquisition indicated that effective provision of corrective feedback can contribute to the development of grammatical competence in second language learners (Fathman and Whalley, 1990; Ashwell, 2000; Ferris and Roberts, 2001; Chandler, 2003). Currently, in the field of natural language processing, the development of tools and technologies to automatically detect grammatical errors is an important research trend.

On the one hand, regarding common error types made by learners of English as a second language and the development of the related automatic detection research, Donahue (2001) used the error taxonomy of native English learners proposed by Connors and Lunsford (1998) to analyze two hundred writing tests taken by learners of English as a second language. The most common error types committed by learners of English as a second language were found to differ from those of native English learners. The three most common error types of learners of English as a second language were incorrect usage of commas, incorrect word usage, and missing words. However, this corpus was insufficient to understand common error types committed by most learners of English as a second language. Cambridge University Press collaborated with the University of Cambridge to create the Cambridge Learner Corpus (CLC), which tags approximately 16 million words. Among these words, the three most common error types are incorrect word selection, preposition errors, and determiner errors (Nicholls, 2003).

During the past ten years, natural language processing specialists have designed automated grammatical error detection techniques and tools focused on common error types in the corpus. Examples include preposition error detection by Eeg-Olofsson and Knuttson (2003), Tetreault and Chodorow (2008), DeFelice and Pulman (2009) and Tetreault and Chodorow (2009), article and preposition error detection by Gamon et al. (2009) and Dale and Kilgarriff (2011), determiner and
proposition error detection by Dale et al. (2012), and determiner, article, and proposition error detection by Ng et al. (2013).

On the other hand, regarding common error types in learners of Chinese as a second language, Wang (2011) observed that the most common grammatical error types among Chinese learners whose mother tongue is English are missing language components, incorrect word order, and incorrect sentence structure. Additionally, analysis of the HSK corpus of 35,884 erroneous sentences has demonstrated that the three most common error types are incorrect word order, missing adverb components, and missing predicate components (Cheng et al. 2014). With the development of related automatic detection research, Cheng et al. (2014) and Yu and Chen (2012) designed word order error detection technology focused on the Chinese sentences in the HSK Dynamic Composition Corpus. In developing a sentence grammatical error detection system, Lee et al. (2014) further used the HSK Dynamic Composition Corpus and additional manually constructed rules of common Chinese sentence errors.

The above literature indicates that, in English learning, there exists widespread use of learning assistance tools developed from natural language processing technology. These tools can automatically detect and correct the grammatical errors of learners. This is valuable as a means to help learners learn correct grammar and improve their compositional skills (Chodorow et al., 2012; Leacock, Chodorow, Gamon, & Tetreault, 2010). However, little research has examined automatic detection of Chinese grammatical errors. This study proposes the integration of rule-based and machine learning methods to identify reliable rules from the corpora to detect the grammatical errors of learners of Chinese as a second language.

3. Corpora

This study seeks to obtain reliable rules to detect grammatical errors committed by learners of Chinese as a second language. The three corpora used in this study include (1) the dry run data provided by the convention; (2) the formal run data provided by the convention; (3) Chinese Written Corpus developed herein. The following focuses on introducing Chinese Written Corpus.

This study has continually developed a Chinese Written Corpus primarily comprising a single topic at different levels. This corpus was developed using Chinese composition scoring guidelines based on the ACTFL (2012) language proficiency criteria. The research samples are compositions written by foreign students who have learned enough Chinese to have basic competence. Samples were collected from September 2010 to June 2013. The source of the corpus is foreign Chinese learners studying at the National Taiwan Normal University Mandarin Training Center and 11 other Taiwanese Chinese educational institutions. The corpus currently includes foreign learners representing 37 different mother languages. During composition collection, complete information was collected on each composition. This information included the title of the composition, the Chinese and English names, nationality, and mother tongue of the learner, and the Chinese education institution in Taiwan. This information was saved as text and image documents. Currently, the texts of this corpus deal with two topics, and there are 1,147 compositions in total, comprising approximately 750 thousand words.

Following the creation of the corpus, each composition text was assessed by two experts or personnel trained in evaluation. To ensure reliability, the texts were cross-evaluated using the Chinese Composition Scoring Standard. This standard assigns compositions to different rankings of Distinguished, Superior, Advanced, Intermediate, and Novice. The Advanced, Intermediate, and Novice categories are each divided into three subcategories, including High, Medium, and Low, amounting for a total of 11 categories. These 11 categories account for Chinese users of all levels, from learners unable to construct a full sentence to native level writers. This study manually scored the compositions in each complete topic based on the above scoring standards and procedures. The composition scores were collected, and learner and corpus information were inputted into an error tagging system developed herein for compositions by learners of Chinese as a second language.

This error tagging system compiles learner and corpus information, and also includes word segmentation, part-of-speech tagging, and error tagging functions. In the main error tagging system, methods and standards for the analysis of learner language errors can be roughly divided into “linguistic form taxonomy” and “surface structure taxonomy”. Linguistic form taxonomy classifies error types – word class, sentence, and specific sentence errors – using language components as a
structure. Meanwhile, surface structure taxonomy classifies error types using their structure. That is, it compares the correct and incorrect forms. Typical surface structures comprise four categories: omission, addition, selection, and disorder (Dulay, Burt & Krashen, 1982; James, 1998). The function of error tagging in this study integrates the two taxonomies, first classifying errors based on the surface structure, then carefully analyzing them based on the language form.

4. Rule Generation and Extraction

Based on the above three data types used in this study, this section explains the optimal method of generating error detection rules to identify ungrammatical sentences.

4.1 Manually Constructed Rules

The study uses five steps to generate manually constructed rules. First, based on the training data provided in this shared task, this study handcrafted syntactic patterns of grammatically incorrect sentences and corrected sentences. Second, to ensure the reliability of manually constructed rules for detecting incorrect sentences, this study also devised a program in which the Chinese Written Corpus developed in this study is embedded. Thirdly, on program completion, we enter syntactic patterns of grammatically incorrect sentences into the interface, and the program can then show the number of sentences contained in the Chinese Written Corpus. Moreover, those sentences conform to syntactic patterns of grammatically incorrect sentences.

Meanwhile, this study entered syntactic patterns of corrected sentences into the program, and then recorded the number of sentences contained in the Chinese Written Corpus, as well as those sentences that conform to the syntactic patterns of corrected sentences. Finally, this study retains the number of syntactic patterns of corrected sentences such that it exceeds that of incorrect sentences. These rules are considered the reliable error detection rules for identifying grammatically incorrect sentences in formal run data. This study contains 840 manually constructed rules, which contain 90 rules for identifying sentences with Missing words, 73 for identifying sentences with Redundant words, 51 for identifying sentences with Selection words, and 626 for identifying sentences with Word disorder.

4.2 Machine Generated Rules

The advantage of manually constructed rules is that complex rules can be detected with high accuracy. However, using manually constructed rules to identify grammatical errors suffers from a disadvantage. Specifically, the number of manually constructed rules is limited, and errors may exist. This study thus employs a program to retrieve syntactic rules of ungrammatical sentences from the learner corpus.

Unlike manually constructed rules, the rules generated by the program are fixed in length. For example, the learner corpus contains the following sentence.

這些地方是在巴西
Neqa Na SHI P Nc

In this sentence, each part of speech is labeled. This sentence in the learner corpus is tagged as the Redundancy error, and ‘SHI’ is a redundant word. We hypothesize that every word in this sentence can be collocated with its beginning and end, and their parts-of-speech to generate rules. Therefore, we combine “是” and its part-of-speech “SHI” with the first and last parts of the word “是” and their associated parts of speech, which yields 32 possible Redundant rules, as shown in Fig. 1.

In Figure 1, the symbol “+” represents two adjacent words or parts-of-speech, while the symbol “>” indicates that both the front and the back of a word or its associated part-of-speech should not be adjacent to that symbol. For a rule pr included in these 32 possible rules, if it meets the following criteria, it will be recognized as an error detection rule:

51
where \( \text{positive}(pr) \) indicates the number of \( pr \) that occurred in the corpus with erroneous sentences; and \( \text{negative}(pr) \) indicates the number of \( pr \) that occurred in the corpus with correct sentence. In this study, the value of \( \text{positive}(pr) \) divided by \( \text{negative}(pr) \) is denoted as the \( r \)-value. The \( r \)-value of rules used by the grammatical error diagnosis algorithm described in Section 5.

Parameters \( p \) and \( k \) are thresholds obtained via experiment. Larger \( p \) is associated with more occurrence of rule \( pr \) in the incorrect sentences. That is, the rule of \( pr \) does not appear randomly. Meanwhile, larger \( k \) represents the possibility of a high degree of precision when using \( pr \) to identify a sentence as erroneous. Take 32 rules in Fig. 1 for example; if \( p \) and \( k \) are set to 2, then just 11 rules with borders in Fig. 1 are collected in the rule base for detection. This study uses the above method to automatically generate 13,890 Redundant rules and 2,497 Missing rules.

\[
\text{positive}(pr) > p \quad \text{and} \quad r > k
\]

\[
r = \frac{\text{positive}(pr)}{\text{negative}(pr)}
\]

Figure 1. Examples of Rules Generated by Machine.

5. Grammatical Error Diagnosis Algorithm

For each sentence, the following steps are performed to determine whether it is incorrect.

Step 1. Check for rules that conform to the error detection rule of Word selection. If such rules exist, the sentence is considered to contain Word selection error and so the error identification is concluded.

Step 2. Check whether rules exist that conforms to the error detection rule of Word disorder. If so, the sentence is considered to contain Word order error and so the error identification is concluded.

Step 3. Check for rules that conform to the error detection rule for Redundant and Missing words.

Step 3.1. If the rule only conforms to one of the error detection rules related to redundant or missing words, then it is considered a sentence that contains that type of error and so the error identification is concluded.

Step 3.2. If the rule simultaneously conforms to more than one error detection rule of redundant or missing words, then among the rules that conforms to both types of error, that with the highest \( r \) value is selected.
Step 3.3. It is assumed that among the Missing word rules, the highest value of $r$ is $mr$, and among the Redundant word rules, the highest value of $r$ is $rr$. If the $r$ value of $rr$ exceeds $y$ times that of $mr$, the sentence is considered to suffer from Redundant word error; otherwise, it is considered to suffer from Missing word error. The identification is concluded following sentence judgment.

Step 4. If the sentence is not recognized as erroneous via the last three steps, then it is considered correct.

Because different types of error detection rules exert different effects, based on analysis of error detection rules from the dry run corpus, their effectiveness reveals that the Selection has higher accuracy than other types of rule. Consequently, when a sentence is identified as containing segments of the rule of Selection, it is recognized that the sentence contains that type of error. Similarly, although the accuracy of the Word disorder rule is lower than that of the Selection rule, it is far higher than that of the Redundant word and Missing word rules. Therefore, when a sentence is identified as containing the Word disorder rule, it is first recognized that the sentence contains that type of error.

Compared to the Missing word rule, the redundant word rule can more easily obtain a higher $r$ value. Thus, if the $r$-value of the Redundant word rule must exceed the missing word rule by $y$ times, then the result of the detection of the rule of Redundant word can be reliable; otherwise, the sentence should be recognized as containing a Missing word error. The next section illustrates the value of each parameter used in the proposed method.

6. Experimental Results

In the NLPTEA 2014 CFL shared task, three parameters are established and combined with three runs to evaluate the effectiveness of the proposed method. In Run 1, the $p$-value is 3, the $k$-value is 2, and the $y$-value is 50. In Run 2, the $p$-value is 10, the $k$-value is 1000 and the $y$-value is 50. In Run 3, the $p$-value is 3, the $k$-value is 2 and the $y$-value is 1. Table 1 lists the experimental results.

Table 1: An example of a table for the ICCE proceedings.

<table>
<thead>
<tr>
<th>Submission Level</th>
<th>Run1</th>
<th>Run2</th>
<th>Run3</th>
</tr>
</thead>
<tbody>
<tr>
<td>False Positive Rate</td>
<td>0.9040</td>
<td>0.2686</td>
<td>0.9040</td>
</tr>
<tr>
<td>Detection Level</td>
<td>Accuracy</td>
<td>0.5006</td>
<td>0.5217</td>
</tr>
<tr>
<td></td>
<td>Precision</td>
<td>0.5003</td>
<td>0.5374</td>
</tr>
<tr>
<td></td>
<td>Recall</td>
<td>0.9051</td>
<td>0.3120</td>
</tr>
<tr>
<td></td>
<td>F1</td>
<td>0.6444</td>
<td>0.3948</td>
</tr>
<tr>
<td>Identification Level</td>
<td>Accuracy</td>
<td>0.2149</td>
<td>0.4109</td>
</tr>
<tr>
<td></td>
<td>Precision</td>
<td>0.2696</td>
<td>0.2516</td>
</tr>
<tr>
<td></td>
<td>Recall</td>
<td>0.3337</td>
<td>0.0903</td>
</tr>
<tr>
<td></td>
<td>F1</td>
<td>0.2983</td>
<td>0.1329</td>
</tr>
</tbody>
</table>

7. Discussion

We have made a few discoveries regarding the process of this experiment and the results obtained. First, manually constructed rules are more complicated than machine-generated rules. However, the accuracy of manually constructed rules does not necessarily exceed that of machine generated rules. Fairly reliable error detection rules can be obtained by establishing parameters based on automatically generated rules. Second, many automatically generated rules are not listed in manually constructed rules. This means the method of using machines to identify error detection rules is feasible. Considering these two perspectives, if the program has an enhanced ability to search for rules, then it is feasible to fully automatically identify grammatical errors made by Chinese as second language learners.
Several aspects of our proposed method can be further improved. First, rules in this study are primarily based on Chinese written error corpus. However, the corpus currently remains in the expansion phase. The increasingly rich content of the corpus can enhance the system performance. Second, only Redundant word and Missing word errors can be automatically generated by the current program. Also, the error detection rules contains only three terms. If more types of rules that are automatically generated by the program can be added in the program and the program can identify more complex rules, the system performance will be further improved.

Acknowledgements

This research is partially supported by the ‘Aim for the Top University Project’ and ‘Center of Learning Technology for Chinese’ of National Taiwan Normal University (NTNU), sponsored by the Ministry of Education, Taiwan, R.O.C. and the “International Research-Intensive Center of Excellence Program” of NTNU and Ministry of Science and Technology, Taiwan, R.O.C. under Grant no. NSC 103-2911-I-003-301

References


Extracting a Chinese Learner Corpus from the Web: Grammatical Error Correction for Learning Chinese as a Foreign Language with Statistical Machine Translation

Yinchen ZHAOa, Mamoru KOMACHIa* & Hiroshi ISHIKAWAa
aGraduate School of System Design, Tokyo Metropolitan University, Japan
*komachi@tmu.ac.jp

Abstract: In this paper, we describe the TMU system for the shared task of Grammatical Error Diagnosis for Learning Chinese as a Foreign Language (CFL) at NLP-TEA1. One of the main obstacles in grammatical error correction for CFL is a data bottleneck problem. The Chinese learner corpus at hand (NTNU learner corpus) contains only 1,208 sentences in total, which is obviously insufficient for supervised learning-based techniques. To overcome this problem, we extract a large-scale Chinese learner corpus from a language exchange site called Lang-8, which results in 95,706 sentences (two million words). We use it as a parallel corpus for a phrase-based statistical machine translation (SMT) system, which translates learner sentences into correct sentences.

Keywords: Chinese learner corpus, web mining, grammatical error correction, statistical machine translation

1. Introduction

Recently, educational applications of natural language processing techniques are actively studied. For example, grammatical error correction for English as a Second Language (ESL) learners has gained large attention in the past few years. Specifically, there were a number of shared tasks of grammatical error correction for ESL learners such as Helping Our Own (HOO) and Conference on Natural Language Learning (CoNLL). However, little attention has been paid to Chinese as a foreign language (CFL). One of the reasons why it is difficult to develop a grammatical error correction system for CFL is the lack of learner corpora. In this paper, we present a method to extract a learner corpus of Chinese from the web, and use it to build a grammatical error correction system for CFL. The main contributions of this paper is as follows:

1. To best of our knowledge, this is the first work that constructs a large-scale learner corpus of Chinese from the web. It contains 100,000 sentences (2M words) annotated with corrections.
2. It is the first work that adopts statistical machine translation (SMT) to grammatical error correction task for CFL. The experimental result shows that our proposed approach is effective to build a precise error correction system.
3. Unlike previous using phrase-based SMT for Chinese spelling correction task, we propose to use character-wise tokenization and prove that character-wise tokenization is more robust than word-wise tokenization.

2. Extracting a Chinese Learner Corpus from the Web

To alleviate the problem of shortage of training data, we resort to extract a Chinese learner corpus from the web. We focus on a language exchange social networking service (SNS)
called Lang-8\textsuperscript{1}. Lang-8 offers a wide variety of languages that you can use to write a blog entry. Other users correct your blog entry written in your learning language, and you in turn correct other users’ blog entry written in your mother tongue. Lang-8 facilitates the process of mutual “language exchange”. Up to date (August 2014), Lang-8 has about one million users where 50,000 of them are Chinese learners.

3. Grammatical Error Correction with Statistical Machine Translation

We decompose the task of grammatical error correction into two parts. First, we identify the location of errors using statistical machine translation trained on a Chinese learner corpus. Second, we classify the type of errors using a simple heuristic rule using dynamic programming.

3.1 Error Identification with Statistical Machine Translation

We follow (Brockett, Dolan, & Gamon, 2006) to make a grammatical error correction system with phrase-based statistical machine translation. One of the advantages of the approach is that we can use an off-the-shelf machine translation toolkit to build a grammatical error correction system if we have a learner corpus with sufficient size.

In their paper, the grammatical error correction process is modeled using a noisy-channel model as follows:

$$\hat{e} = \arg \max_e P(e | f) = \arg \max_e P(f | e)P(e)$$

where $P(e)$ is a language model and $P(f | e)$ is a translation model. In this paper, $f$ corresponds to a learner sentence and $e$ corresponds to a corrected sentence, respectively. The phrase-based SMT toolkit we use in this paper actually uses a log linear model which contains the noisy-channel model as follows:

$$\hat{e} = \arg \max_e w^T h$$

where $w$ is a weight vector and $h$ is a feature function, respectively.

We propose two types SMT systems: word-based system and character-based system, depending on the pre-processing step of a learner corpus. The intuition behind using a character-wise segmentation is that learners of Chinese tend to write incorrect sentences, which may hurt the accuracy of the word segmentation. Character-based SMT is free from tokenization errors, while it is able to learn word-to-word or phrase-to-phrase correction patterns thanks to the phrase extraction heuristics.

3.2 Error Classification with Dynamic Programming

Once we identify the location of errors, we classify the type of errors using a simple heuristic rule. We use a dynamic programming algorithm to calculate the number of insertion, deletion and replacement operations for each sentence pair. We then classify the type of errors by the following pseudo-code:

\texttt{1 http://lang-8.com/}
Table 1: Pseudo-code for error type classification.

<table>
<thead>
<tr>
<th>Input: learner sentence $l$, system correction $c$</th>
<th>Output: error type $t$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$(i, d, r)$ ← get_operations($l, c$)</td>
<td>$(i, d, r)$ ← get_operations($l, c$)</td>
</tr>
<tr>
<td>if $d &gt; 0$ and $i &gt; 0$</td>
<td>if $d &gt; 0$ and $i &gt; 0$</td>
</tr>
<tr>
<td>$t$ ← “Disorder”</td>
<td>$t$ ← “Disorder”</td>
</tr>
<tr>
<td>else if $r &gt; 0$</td>
<td>else if $r &gt; 0$</td>
</tr>
<tr>
<td>$t$ ← “Selection”</td>
<td>$t$ ← “Selection”</td>
</tr>
<tr>
<td>else if $d &gt; 0$</td>
<td>else if $d &gt; 0$</td>
</tr>
<tr>
<td>$t$ ← “Redundant”</td>
<td>$t$ ← “Redundant”</td>
</tr>
<tr>
<td>else if $i &gt; 0$</td>
<td>else if $i &gt; 0$</td>
</tr>
<tr>
<td>$t$ ← “Missing”</td>
<td>$t$ ← “Missing”</td>
</tr>
<tr>
<td>else</td>
<td>else</td>
</tr>
<tr>
<td>$t$ ← “correct”</td>
<td>$t$ ← “correct”</td>
</tr>
<tr>
<td>end if</td>
<td>end if</td>
</tr>
<tr>
<td>return $t$</td>
<td>return $t$</td>
</tr>
</tbody>
</table>

If a sentence contains only one error, this algorithm correctly returns the "Disorder" error type, while it may fail to classify "Selection" error type and output "Redundant" or "Missing" error types. In a preliminary experiment, we found that this confusion can be negligible. We did not explore the use of machine learning-based classification method because the training corpus provided by the organizer contains only 1,000 instances.

4. Experiments

In this section, we describe the experimental settings and results for the NLP-TEA1.

4.1 Data and Tools

We obtained the Lang-8 Learner Corpora v2.0. The corpora come with “blog id”, “sentence id”, “learning language”, “native language”, “learner sentences” and “corrected sentences”. We extracted blog entries whose “learning language” is set to “Mandarin”. The Chinese portion of the Lang-8 Learner Corpora consists of 29,595 blog entries (441,670 sentences). We discarded following sentences and kept 95,706 sentences at last.

- Too long (more than or equal to 20 words) or too short (less than or equal to 3 words).
- Not written in Chinese.
- Any corrected sentence 1.3 times longer or more than the original one.\(^2\)

We used Moses 2.1.1 as a statistical machine translation toolkit with its default parameter. The training and testing was done using the scripts distributed as KFTT Moses Baseline v1.4 (Neubig, 2011). We did not perform minimum error rate training (Och, 2003). We trained an SMT system with two training corpora: the Lang-8 Chinese Learner Corpus with and without word segmentation. In other words, we built a grammatical error correction

\(^2\) Some corrected sentences contain comments and annotations, which may harm word alignment for SMT.
system trained on a character-based phrasal SMT in addition to a word-based phrasal SMT. We used jieba\textsuperscript{3} 0.32 for Chinese text segmentation.

4.2 Results

Table 1 summarizes the false positive rate, accuracy, precision, recall and F1 scores for the formal run. Character-based approach outperformed word-based approach in all evaluation metrics. This confirms the hypothesis that word segmentation errors damage grammatical error correction for CFL.

We ranked the 2\textsuperscript{nd} at the false positive rate and accuracy out of six groups participated in the shared task. However, these evaluation metrics alone cannot verify the effectiveness of our approach, since there is a trade-off between these metrics. Note that we only reports the scores at detection level, since the performance at identification level is almost the same.

Table 1: Experimental results for the formal run at NLP-TEA1. Accuracy, precision, recall and F1 scores are at the detection level.

<table>
<thead>
<tr>
<th></th>
<th>False Positive Rate</th>
<th>Accuracy</th>
<th>Precision</th>
<th>Recall</th>
<th>F1</th>
</tr>
</thead>
<tbody>
<tr>
<td>TMU-Run1:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Character-based</td>
<td>0.1977</td>
<td>0.5171</td>
<td>0.5399</td>
<td>0.2320</td>
<td>0.3245</td>
</tr>
<tr>
<td>TMU-Run2:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Word-based</td>
<td>0.1691</td>
<td>0.5103</td>
<td>0.5287</td>
<td>0.1897</td>
<td>0.2792</td>
</tr>
</tbody>
</table>

5. Discussion

Our system achieved the worst (6/6) performance in terms of F1 score. The main reason is that we did not perform any parameter tuning at all, even though the error distribution of the test corpus is very skewed (half of the sentences contain errors). In a preliminary experiment, we ran the minimum error rate training using BLEU (Papineni, Roukos, Ward, & Zhu, 2002), but after the optimization the system outputs almost no corrections. This is because the BLEU score will become higher if the system does not change the learner sentence. Although BLEU is used to evaluate grammatical error correction as in (Park & Levy, 2011), it may not adequate to assess the quality of error correction systems. One possible direction is to optimize the SMT system using the F1 score with Z-MERT\textsuperscript{4}.

Note that the shared task only requires participants to determine whether a given sentence contains an error or not, our system is capable of locating the position of errors. In addition, our system can identify multiple errors in a sentence (although it is out of scope of this shared task).

One of the side effects of using the Lang-8 corpus is that the error correction system misclassifies correct sentences as "Missing" errors since it tends to use commas where applicable. However, commas often make more natural Chinese expressions than original. For instance, consider the following example. The system output is more fluent than the original, but it is different from the gold standard annotation, which deteriorates performance.

\textbf{Gold}: 今天的天氣很好不怎麽熱 （Today's weather is good, not very hot.）

\textbf{System}: 今天的天氣很好，不怎麼熱

\textsuperscript{3}https://github.com/fxsjy/jieba
\textsuperscript{4}http://cs.jhu.edu/~ozaidan/zmert/
Also, we would like to emphasize that we did not use any resources provided by the organizer. It is interesting to use domain adaptation approach such as in (Imamura, Saito, Sadamitsu & Nishikawa, 2012) to better reflect error distribution of the given domain (for example, 50% of the given test corpus contains errors, which is not often the case in realistic setting).

If a sentence contains more than one error, the proposed error type classification algorithm will output only one error type. Since the test corpus is controlled to contain only one error, we opted for a simple rule for the shared task. However, it is possible that these error types are not identical in real setting, so our future work includes error type classification for each error.

6. Related Work

Lang-8 is considered as one of the invaluable resources for knowledge acquisition for second language learners. For example, Japanese learner corpus (Mizumoto, Komachi, Nagata, & Matsumoto, 2011; Kasahara, Komachi, Nagata, & Matsumoto, 2011) and English learner corpus (Tajiri, Komachi, & Matsumoto, 2012; Mizumoto, Hayashibe, Komachi, Nagata, & Matsumoto, 2012) can be extracted from Lang-8. It is not surprising that we can extract a large corpus of Chinese learners since Chinese (Mandarin) is the third most popular learning languages in Lang-8, followed by English and Japanese.

The use of statistical machine translation techniques to grammatical error correction was pioneered by (Brockett, Dolan, & Gamon, 2006), and has been adopted to many researchers in grammatical error correction for ESL (Mizumoto, Hayashibe, Komachi, Nagata, & Matsumoto, 2012; Buys & van der Merwe, 2013; Yuan & Felice, 2013; Behera & Bhattacharyya, 2013; Junczys-Dowmunt & Grundkiewicz, 2014).

Recently, similar approach is applied to Chinese spelling error correction as well (Wu, Liu & Lee, 2013; Wu, Chiu & Chang, 2013; Liu, Cheng, Luo, Duh & Matsumoto, 2013). However, all of these methods use word-based statistical machine translation, even though some of them use character n-gram language model. One of our proposed models investigates character-wise segmentation rather than word-wise one, and indicates that character-based model can learn useful correction patterns if the training corpus is sufficiently large.

Error type classification has gained much attention, for example in English (Swanson & Yamangil, 2012) and Japanese (Oyama, Komachi & Matsumoto, 2013). Although these works use linguistically motivated annotation scheme proposed in previous work, the error type annotation scheme for the NTNU learner corpus is based on edit operations and it is more appropriate to use rules rather than machine learning.

7. Conclusion

In this paper, we described the TMU system for the Grammatical Error Diagnosis for CFL Shared Task at NLP-TEA1. To increase the number of training corpus, we explored the web for constructing a learner corpus of Chinese. We extracted 100,000 learner sentences paired with their correction from the language exchange SNS, Lang-8, and used it to train an SMT-based grammatical error correction system. We compared two types of segmentation for phrasal SMT and found that character-based SMT outperforms word-based SMT for CFL grammatical error correction. The system achieved moderate performance even though it did not use any language resources from the target domain.

5 http://cl.naist.jp/nldata/lang-8/
Acknowledgements

We would like to thank Xi Yangyang for granting use of extracted texts from Lang-8.

References


Detecting Grammatical Error in Chinese Sentence for Foreign

Jui-Feng Yeh, Yun-Yun Lu, Chen-Hsien Lee, Yu-Hsiang Yu, Yong-Ting Chen
Department of Computer Science and Information Engineering, National Chiayi University
No.300 Syuefu Rd., Chiayi City 60004, Taiwan (R.O.C.)
{ralph, s1020447, s1020443, s1002967, s1003008}@mail.ncyu.edu.tw

Abstract: Each language has its own unique grammar, Chinese is same as the other languages similarly. But each language has different type even there does not have any relations. So the foreigners learn the language not only need to learning the word pronouns and glyph, but also need to learn the grammar. This issue is very extensive, not only can help foreigners to learn Chinese, but also can detect the error grammar. This paper had proposed method can divide five sections of the structure: First sections are input sentence; second sections are parsed and word segmentation; third sections find the missing, wrong word; fourth sections find the redundant wrong word; fifth sections are final output. This paper has two parts, the first is how to detect the grammar error, and the second is how should we know the Chinese grammar error is what type. Finally, we can get the type of the grammar, and we can know how to correct.

Keywords: CFL, Chinese word correction, grammatical error, rule induction.

1. Introduction

Learning Chinese is more famous than before, not only there has more and more Chinese business, but also there has many tourist attractions cause the foreigners love to go there to travel. So, there have many of foreigners beginning to learn Chinese. But Chinese is not easy to learn like the other languages, it is not only having many pronunciations and glyph of the word, but also have many grammar of the sentence. The Chinese although has subject, verbs and object too, but there has a combination of fixed text, if you do not to comply with these rules, the meaning of the sentence itself will be different. So how to learn Chinese grammar is very important research. This topic is extensive, not only for foreigners to learn Chinese, but also can help to detect the wrong grammar in the document.

In recent years, there has a lot of paper to research about Chinese learning grammar. Most of paper about learning Chinese paper not only talk about the sentence correct rate, but also talk about the Grammar correct rate. Ying Jiang (2012) proposes an arithmetic called “Rnture-Sentence” to segment the Chinese sentences, it can segment the sentence more completely, and solve the problem about the complicated Chinese grammar, proofreading method, their method to cross sentences is based on LanguageTool, their paper also presents to some method of new rule which can accomplish complicated Chinese grammar proofreading. These authors also propose another Chinese grammar, proofreading (2013), the presents an indexing method of a corpus base of the Chinese grammar, they can evaluate the rule of the accuracy and the frequency, each rule, they adopt an iterative approach to improve it, make sure its better performance in the real word, it also introduces the important role of the corpus of their method. Mei-Jen Audrey Shih et al. (2011) propose a Chinese online learning system, this online system is convenience and it is assembled to abound environment and had a broad content search opportunity, this paper is mained on how to learn Chinese language effectively in an online learning environment. Lee Jo Kim et al. (2011) propose a Chinese language teaching and learning system based on ICT-Base tool, this tool can help peer assisted learning environment. Ying Jiang et al. (2012) they provide a new method to deal with grammar error. They arrange a new rule for their new grammar system, this system has two combined characters with intention, then their grammar system can contain as well as the spelling error system, so their system can have perfect respect of precision and practicality. David Tawei Ku et al. (2012) proposes a situated learning for
Chinese learning, it is a trend that ubiquitous learning environment, and the feature main on real life learning situation and problem solving practice, this learning system has two parts, one is integrating situated learning strategy and the other is context awareness technology. Yanwei Wang et al. (2011) proposes a discriminative learning method of MQDF (Modified Quadratic Discriminant Function), MQDF is based on sample importance weights, this method is investigated and compared other discriminative learning methods about MQDF. Lung-Hsiang Wong (2010) propose a Mobile-Assisted Language Learning (MALL), there have two case studies, and mained on "creative learner outputs", student in two studies language by one-to-one mobile devices, and capture the picture of the real life. Hui Yang et al. (2010) proposes a continuous prior polarity algorithm, their method reflects subtle changes of sentiment in contrast, its previous studies which expressed sentiment polarity discretely, they also proposed a method based on Chinese dependency grammar which can assess modified polarity, they can accurately identify subjective words and its modified according different Chinese dependency grammar, then predict the sentence by aggregate. Peng Li et al. (2012) proposes “A Hierarchy-based Constraint Dependency Grammar Parsing for Chinese”, they mentioned the Constraint Dependency Grammar (CDG) is a famous formalism which about the grammatical rules, and they have successfully adopted in Chinese, they propose to develop a three schema which is based on a study of constraining in the corpus. Haiping Zhu et al. (2011) propose a analyze Chinese sentence with semantic dependency method, the correlation between words and phrases can calculation of the similarity from Hownet, between the sentence and sentence’s similarity can analyze by formula, their method can be adopted to analyze the correct topic and to categorize.

This paper had proposed method can divide five sections of structure: First sections are input sentence; second sections are parsing and word segmentation; third sections are find the missing wrong word; forth sections are find the redundant wrong word; fifth sections are final output. In the third section, we classify the four of the grammar rule, the part of speech(POS) can classify four type (Shi, Neu, D, DA), and there have regular POS behind the four type of the POS. After find, we put the wrong grammar part in the dictionary file which only for these error, the dictionary file name is Miss. In the forth section, find the redundant wrong word, there has a special rule behind the POS of DE must a DE, if did not, it will write in the dictionary file which name is “Redundant”. finally, we will use these dictionary files to detect the Chinese Grammar, if the error is bellowing “Missing” dictionary file, the error is bellowing “Redundant” dictionary file, if not miss error or redundant error, that means it it correct, output to correct the file.

2. Method

![Figure 1. The framework of the proposed system.](image)

In this section, we will introduce the framework of the proposed system and method. Our proposed method is aimed to detect and identify the sentence for learning Chinese as a foreign language (CFL). The sentences written by CFL may contain a variety of grammatical errors, such as word choice, missing words, word disorder and so on. It focuses on grammatical errors in this task. And the
framework is divided into two parts: training phase and test phase that will describe in section 2.1 and section 2.2.

2.1 Training phase

The training phase shown in figure 1, we have some data that are contain some sentences can be trained to find some useful information. First, we do the pre-process to the data from the task organizer, that will be input file and we removed unneeded portions of each sentence in this file, such as SID number, then the treated results will be further inputted into the tool which is CKIP AutoTag, that is to do the word segmentation and part-of-speech (POS) tagging based on E-Hownet. The corresponding part-of-speech of each word is obtained in the sentences, which is given a part of speech at the end of a word in parentheses. Second, we are going to remove unessential blank spaces and parentheses, that will be more convenient in the following file operations. In the test phase, we are also adopted in this way. Then, we want to find some rules with training data which can be used in test phase. We construct the training data rule from the results of process which have part-of-speech. Finally, the candidate outputs are generated according to our training data rule.

2.2 Test phase

In the previous section, the training data rule is built in training phase. We will describe the test phase of the framework in this section. The word segmentation and part of speech (POS) labeling are the same as training phase. Then, we begin the processes with the third & fourth step, we have to detect and identify the wrong word. The following is focused on finding the Missing type of the wrong words.

- Behind the word with POS of “Shi” is not connected the word with POS of “Verb”.
- Behind the word with POS of “Neu” is not connected the word with POS of “De”.
- Behind the word with POS of “D” is not connected the word with POS of “Neqa”.
- Behind the word with POS of “Da” is not connected the word with POS of “Neu”.

According to above, the Missing type of the incorrect words will save in the file named missing. And the following is focused on finding the Redundant type of the wrong words.

- Behind the word with POS of “De” is not connected the word with POS of “De”.
- Behind the word with POS of “P” is not connected the word with POS of “P”.
- Behind the word with POS of “Cbb” is not connected the word with POS of “D”.
- Behind the word with POS of “Vh” is not connected the word with POS of “D”.
- Behind the word with POS of “De” is not connected the word with POS of “Neqa”.
- Behind the word with POS of “D” is not connected the word with POS of “D”.

According to above, the Missing type of the incorrect words will save in the file named missing. We will remove repeated SID. It is helping us to reduce the process time. We will output the result in the final step. The processes will run the test data, if the word If the word exists in the missing file, we will output the sentence with Missing. The Redundant type of the incorrect word is same as Missing. Others are identified as correct. For example, input: “(sid=C1-1876-2) 對社會國家不同的影響”, output: “C1-1876-2, Missing”, If the input contains no errors, the system should return “C1-1876-2, correct”.

3. Experiments

According to the grammatical error diagnosis for learning Chinese as a foreign language in NLP-TEA-1, this paper is dedicated to the detection and identification of errors in sentences. The evaluate is divided into two parts: Subtask 1 is detection level that is to check out the sentence which
is incorrect or correct, then the subtask 2 is identification level, which is to identify the error type in sentences, i.e., Redundant, Missing, Disorder, and Selection. In section 3.1, we will describe the data sets, performance metrics, then we will show our evaluation in section 3.2.

3.1 Data sets

![Figure 2. An example of the training data.](image)

In this task, the evaluation is an open test. Participants can employ any linguistic and computational resources to develop the error diagnosis, and provide data of CFL’s essays from the NTNU learner corpus for training purpose. The corpus was released in SGML format which is shown in figure 2. Moreover, there are at least 1000 different degrees of difficulty of testing passages for testing. In this paper, we use C++ to develop our proposed method.

![Figure 3. A quadrant map of performance metrics.](image)

The judging correctness are divided into two parts: detection level and identification level. The following are showing some performance metrics and quadrant map shown in figure 3 that is measured in both levels of indicators:

- **TP**: System determines the character for errors related to the actual error, and the judgments the system is correct.
- **FP**: System determines the character for errors is not related to the actual error, and the judgments of the system is incorrect.
- **FN**: System determines the character for errors is related to the actual error, and the judgments of the system is incorrect.
- **TN**: System determines the character for errors is not related to the actual error, and the judgments of the system is correct.

The following of performance metrics are according to the quadrant map.

- **False Positive Rate** = \( \frac{FP}{FP+TN} \)
• **Accuracy** = \( \frac{TP+TN}{(TP+TN+FP+FN)} \)

• **Precision** = \( \frac{TP}{TP+FP} \)

• **Recall** = \( \frac{TP}{(TP+FN)} \)

• **F1 Score** = \( \frac{2 \times \text{Precision} \times \text{Recall}}{(\text{Precision} + \text{Recall})} \)

### 3.2 Evaluation

According to the table 1, our false positive rate is the best in this task, which means that our proposed method is feasible, but our proposed method just focuses on identifying two error type,. There are two parts of performance evaluation: detection level and identification level which is shown in table 2 and table 3. In the identification, we can see that accuracy is the best. Then, accuracy and precision are also comparable to others, but our method in recall is relatively weaker than another. This performance evaluation shows that our method is viable, but our method is still much room for improvement.

<table>
<thead>
<tr>
<th>Participating teams</th>
<th>False Positive Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>NCYU*</td>
<td>0.1189</td>
</tr>
<tr>
<td>TMU</td>
<td>0.1691</td>
</tr>
<tr>
<td>UDS</td>
<td>0.2491</td>
</tr>
<tr>
<td>KUAS&amp;NTNU</td>
<td>0.2686</td>
</tr>
<tr>
<td>CIRU</td>
<td>0.496</td>
</tr>
<tr>
<td>NTOU</td>
<td>1</td>
</tr>
</tbody>
</table>

**Table 2.** Participating teams of performance evaluation in Detection Level.

<table>
<thead>
<tr>
<th>Participating teams</th>
<th>Accuracy</th>
<th>Precision</th>
<th>Recall</th>
<th>F1</th>
</tr>
</thead>
<tbody>
<tr>
<td>NCYU*</td>
<td>0.4983</td>
<td>0.4927</td>
<td>0.1154</td>
<td>0.187</td>
</tr>
<tr>
<td>TMU</td>
<td>0.5171</td>
<td>0.5399</td>
<td>0.232</td>
<td>0.3245</td>
</tr>
<tr>
<td>UDS</td>
<td>0.4914</td>
<td>0.4945</td>
<td>0.7749</td>
<td>0.6037</td>
</tr>
<tr>
<td>KUAS&amp;NTNU</td>
<td>0.5006</td>
<td>0.5003</td>
<td>0.9051</td>
<td>0.6444</td>
</tr>
<tr>
<td>CIRU</td>
<td>0.6446</td>
<td>0.6128</td>
<td>0.7851</td>
<td>0.6884</td>
</tr>
<tr>
<td>NTOU</td>
<td>0.5</td>
<td>0.5</td>
<td>1</td>
<td>0.6667</td>
</tr>
</tbody>
</table>

**Table 3.** Participating teams of performance evaluation in Identification Level.

<table>
<thead>
<tr>
<th>Participating teams</th>
<th>Accuracy</th>
<th>Precision</th>
<th>Recall</th>
<th>F1</th>
</tr>
</thead>
<tbody>
<tr>
<td>NCYU*</td>
<td>0.4594</td>
<td>0.2409</td>
<td>0.0377</td>
<td>0.0652</td>
</tr>
<tr>
<td>TMU</td>
<td>0.4554</td>
<td>0.3545</td>
<td>0.1086</td>
<td>0.1662</td>
</tr>
<tr>
<td>UDS</td>
<td>0.2337</td>
<td>0.2467</td>
<td>0.2594</td>
<td>0.2529</td>
</tr>
<tr>
<td>KUAS&amp;NTNU</td>
<td>0.2149</td>
<td>0.2696</td>
<td>0.3337</td>
<td>0.2983</td>
</tr>
<tr>
<td>CIRU</td>
<td>0.4589</td>
<td>0.4548</td>
<td>0.4137</td>
<td>0.4333</td>
</tr>
<tr>
<td>NTOU</td>
<td>0.2074</td>
<td>0.2932</td>
<td>0.4149</td>
<td>0.3436</td>
</tr>
</tbody>
</table>

### 4. Conclusions

This study proposes a method for Chinese text detect grammar error. The method in our study is focus on word classify to easy detect Chinese grammar error. The grammar error is classifying four type, the verbs was not add behind POS of Shi, the De was not add behind the POS of Neu, the Neqa was not add behind POS of D, and the Neu was not add behind POS of Da. The experimental result shows the
performance it good, and we also apply this method in “grammatical error diagnosis for learning Chinese as a foreign language”, and the final result pretty good. In the feature, we hope can raise the performance and find the more grammar type. More grammar type can helpful to find the Chinese grammar error. After the Chinese grammar error, we will start to study the relationship between grammar and spelling errors, because in this paper we only care about the word pronouns and glyph, but in recent years some spelling error has been regularization, it most to understanding the context then detect it is right or wrong, so the issue about the relationship between grammar and spelling errors is need to study, if we can fine the relationship then the Chinese grammar detect correct rate must can raise higher.

Acknowledgements

This work is supported in part by the National Science Council, Taiwan, R.O.C., under the project grant numbers NSC 102-2221-E-415-006-MY3.

References


Academia Sinica CKIP. http://ckipsvr.iis.sinica.edu.tw/

Grammatical Error Detection with Limited Training Data: The Case of Chinese

Marcos ZAMPIERI, Liling TAN
Saarland University, Germany

Abstract: In this paper, we describe the UDS submission to the shared task on Grammatical Error Diagnosis for Learning Chinese as a Foreign Language. We designed four different experiments (runs) to approach this task. All of them are variations of a frequency-based approach using a journalistic corpus as standard corpus and comparing n-gram frequency lists to both the training and the test corpus provided by the shared task organizers. The assumption behind this approach is that comparing a standard reference corpus to a non-standard study corpus using frequency-based methods levels out non-standard features present in the study corpus. These features are very likely to be, in the case of this corpus, grammatical errors. Our system obtained 60.3% f-measure at the error detection level and 25.3% f-measure at the error diagnosis level.

Keywords: grammatical error detection, Chinese, error diagnosis, learner corpora

1. Introduction

Grammatical error detection and correction is a vibrant research area in NLP. In the last couple of years much effort has been concentrated on the detection, diagnosis and correction of errors in texts written both by native speakers and by foreign language learners. For foreign language learning the practical applications of grammatical error detection are manifold, ranging from spelling and grammar checkers to essay scoring and grading.

Given this interest, a number of shared tasks have been organized in recent years. This includes the HOO 2012 preposition and determiner error correction shared task (Dale et al., 2012) held at the 2012 edition of the BEA Workshop and the Grammatical Error Correction shared tasks held at CoNLL-2013 (Ng et al., 2013) and one year later at CoNLL-2014 (Ng et al., 2014).

Similar to the previous shared tasks, this year’s Grammatical Error Diagnosis for Learning Chinese as a Foreign Language provided us the opportunity to explore computational methods on diagnosis of errors committed by foreign learners of Mandarin Chinese. The shared task was designed to evaluate systems’ output in two stages:

1) Error detection level: identify whether a sentence contains an error or not.
2) Error diagnosis level: classification of errors types (redundant words, missing words, word order and word selection).

Participants were required to train their systems not only to identify errors, but also to classify error types making the task more challenging. As an example, a system trained for Chinese error diagnosis was recently presented by Lee et al. (2014) obtaining 68.9% F1 score.

Apart from the difficulty of the task itself, it is important to note that the computational processing of logographic languages such as Chinese poses several difficulties to researchers used to handling character-based languages. Trivial pre-processing tasks like tokenization and segmentation are much more challenging for Chinese than for example, for English. This issue will be discussed in more detail in this paper.
In the next sections we describe the UDS submissions to the shared task commenting on the results obtained and on the strengths and weaknesses of our approach. In our submissions we used a frequency-based approach using a reference corpus to compensate the small amount of training data available.

2. Related Work

Grammatical error correction and detection has been the subject of a number of research papers in recent years. Shared tasks such as the aforementioned Grammatical Error Correction at CoNLL-2013 (Ng et al., 2013) and CoNLL-2014 (Ng et al., 2014) have been organized to evaluate systems' performance in correcting errors in learner corpora.

Tetreault and Chodorow (2008) presented experiments for detecting preposition errors in English texts written by non-native speakers. The authors report 84% precision and 19% recall. Heilman et al. (2012) proposed a hybrid error correction approach to the HOO 2012 shared task (Dale et al., 2012) focusing on increasing recall and F-measure scores. The authors argue that most systems take only precision into account due to the high cost of false positives (e.g. labeling grammatical sentences as ungrammatical).

More recently, Yuan and Felice (2013) proposed the use of phrase-based statistical machine translation to grammatical error correction. The application of SMT techniques to the task is not new (Brockett et al., 2006) and the performance achieved by their approach is not particularly high. However, in Yuan and Felice (2013), authors contribute in other ways, as for example, in exploring methods of generating new artificial errors to increase the size of the dataset and therefore providing more training material. The generation of artificial errors has been the subject of other research papers such as in Foster and Anderson (2009) and Felice and Yuan (2014).

As for Chinese, Yu and Chen (2012) investigated the problem of word ordering errors in Chinese texts written by Chinese foreign language learners. Authors report 71.64% accuracy using word n-grams and POS tags. Chang et al. (2012) presented a rule-based learning algorithm (first order inductive learner (FOIL)) combined with a log-likelihood function to identify error types in Chinese texts.

In this section we briefly discussed a couple of recent papers that deal with error detection, correction and diagnosis. For a comprehensive overview about the topic see Leacock et al. (2014).

3. Methods

Given the task description presented in section 1, we received a training corpus from the organizers containing over 12,000 labeled instances in XML format. The corpus was annotated with a unique identifier for each sentence 'sentence id', the type of mistake that each sentence contained and its respective correction. A snapshot of the corpus provided by the organizers can be seen next:

```xml
<ESSAY title="寫給即將初次見面的筆友的一封信">
  <TEXT>
    <SENTENCE id="B1-0112-1">我的計畫是十點早上在古亭捷運站</SENTENCE>
    <SENTENCE id="B1-0112-2">頭會戴著藍色的帽子</SENTENCE>
  </TEXT>
  <MISTAKE id="B1-0112-1">
    <TYPE>Disorder</TYPE>
    <CORRECTION>我的計畫是早上十點在古亭捷運站</CORRECTION>
  </MISTAKE>
  <MISTAKE id="B1-0112-2">
    <TYPE>Missing</TYPE>
    <CORRECTION>頭上會戴著藍色的帽子</CORRECTION>
  </MISTAKE>
</ESSAY>
```

In our preliminary experiments we observed that the corpus provided was not sufficiently large to build robust machine learning models for grammatical error detection or diagnosis. In a similar text classification shared task using learner corpora (Tetreault et al., 2013), the amount of training data available was not sufficient to train effective models.
was significantly larger than the test data which allowed researchers to build more robust models based only on the given training data.

In addition to that, we had a couple of problems with the Chinese segmentation tool that we used (Chang et al. 2008) and this returned us fewer segments than were actually in the training corpus. We unfortunately did not have enough time to perform error analysis on the segmentation and pre-processing tools available nor did we have time to use the most recent Chinese segmenters (Tan and Bond, 2014; Wang et al., 2014) before the shared task submission deadline. Given these difficulties, we had to search for new strategies to approach the task with limited training data that could still achieve results comparable to the state-of-the-art systems. Inspired by existing related work, we considered three alternatives to approach the task.

a) Use an external Mandarin Chinese as a foreign language corpus preferably containing similar tags to those of the training and test data.

b) Generate a list of artificial errors to increase the amount of instances in the training corpus as in Felice and Yuan (2014).

c) Use a frequency-based approach to compare the learner corpus to a standard general language corpus. The assumption is that this comparison would level out non-standard features of the training/test data that are more likely to be errors.

Given the time and resources we had, we decided to go with option (c) and leave the other two for future work. Option (a) seemed to be promising and straightforward in terms of performance, but we did not have suitable training data at our disposal. Acquiring and annotating new data is expensive and time consuming which made option (a) infeasible. As to option (b) we regard it to be a suitable and interesting alternative in cases where training data is not available. However, it is not currently possible to say much about the performance of these methods for Chinese. To our knowledge, previous work has only been done for Indo-European languages.

Option (c) proved to be the most adequate solution for our submission. A frequency-based approach, like the one used in our submission, requires only a large reference corpus (a general standard contemporary language one). We had a couple of suitable resources at our disposal (Chen et al., 1996 Graff and Chen, 2003) and we therefore decided to test this method.

The method works under a similar assumption to the keyword lists widely used in corpus linguistics (Scott, 1997; McEnery, 2009) and also applied on a similar scenario by Zampieri et al. (2013) on Internet data. Keyword lists are produced by comparing two corpora (a study corpus and a reference corpus) using association metrics such as log-likelihood, chi-square or mutual information. These keywords usually reflect salient features of the study corpus. In the case of the present comparison (standard corpus versus learner corpus), it is safe to assume that a reasonable amount of salient features from the learner corpus will be infrequent distributions of words which are very likely to be errors. This is the basic assumption of our approach.

3.1 Algorithm

If one assumes that a reference corpus is a portrait of standard language, lexical items that stand out in the study corpus in comparison to the reference corpus should deviate from what is considered to be 'the norm'. This is a relatively naive assumption and thematic bias may still occur when using unbalanced data. To avoid that, we used a large balanced journalistic corpus (Graff and Chen, 2003) as our standard corpus. From the reference corpus we sampled the first 50,000 sentences and extracted n-grams (1 to 5) using the KenLM Language Model Toolkit (Heafield, 2011).

We pre-processed the training, test and standard corpora using the Stanford tokenizer (Chang et al. 2008). As Chinese is a logographic language we treat every character in isolation. As previously mentioned, the Stanford segmenter yielded a number of errors in segmentation that worsened our system's performance. However, we were not able to evaluate the exact segmenter's performance for our dataset before this submission.

From the training and test corpora provided by the shared task organizers we proceeded to extract a list of ungrammatical n-grams that were not present in the subset of the reference corpus and treated them as key expressions. This calculation returned us a list of 35,000 ungrammatical n-grams not present in the reference corpus.

It is important to note here that the main difference between our approach and what is commonly used in corpus linguistics is that the latter uses the lexicon in the form of bag-of-words (or
less often bigrams). In these experiments we used the complete set of n-grams (1 to 5) extracted from the corpus thus increasing the coverage of our method.

With these n-gram lists, we trained two classifiers to identify grammatical and ungrammatical instances: 1) a simple n-gram-based classifier to identify correct (grammatical) sentences using the formula below and 2) a Multinomial Naive Bayes (MNB) classifier to identify ungrammatical sentences along with their labels using the Scikit-learn package (Pedregosa et al., 2011).

$$\frac{\sum p(n_g | n_g \in n_{\text{gram}}, n_{\text{runout}})}{\sum p(n_g | n_g \in n_{\text{gram}})} > X$$

In the formula above, we tuned the $X$ parameter value to optimize the results obtained by the first classifier. After a number of tests we found that the optimal value lies between 0.10 and 0.20. We therefore produced four submissions (runs) using different X values: 0.20 for the 1st run, 0.16 for the 2nd run, 0.15 for the 3rd run and 0.10 for the 4th run. The best results were obtained in our first run, using $X = 0.20$ and these are the results that will be reported and discussed next.

4. Results

According to the information provided by the organizers, 13 teams registered for the shared task and 6 of them submitted their final results. The results were calculated using standard metrics in text classification, namely: precision, recall, accuracy and F-measure as well as a false positive rate score. No limitation regarding the number of runs was set. The test set provided by the organizers contained 1,750 unlabeled test instances.

The UDS team submitted four runs changing the $X$ parameter of our correct sentence classifier as explained in the previous section. In table 1 we present the best results obtained by all 6 groups at the error detection level. At this level, our approach was the fourth best with results reaching 60.37% F1 score and 49.14% accuracy.

<table>
<thead>
<tr>
<th>Team</th>
<th>Accuracy</th>
<th>Precision</th>
<th>Recall</th>
<th>F1</th>
</tr>
</thead>
<tbody>
<tr>
<td>CIRU</td>
<td>0.6446</td>
<td>0.6128</td>
<td>0.7851</td>
<td>0.6884</td>
</tr>
<tr>
<td>NTOU</td>
<td>0.5000</td>
<td>0.5000</td>
<td>1</td>
<td>0.6667</td>
</tr>
<tr>
<td>KUAS&amp;NTNU</td>
<td>0.5006</td>
<td>0.5003</td>
<td>0.9051</td>
<td>0.6444</td>
</tr>
<tr>
<td><strong>UDS</strong></td>
<td><strong>0.4914</strong></td>
<td><strong>0.4945</strong></td>
<td><strong>0.7749</strong></td>
<td><strong>0.6037</strong></td>
</tr>
<tr>
<td>TMU</td>
<td>0.5171</td>
<td>0.5399</td>
<td>0.232</td>
<td>0.3245</td>
</tr>
<tr>
<td>NCYU</td>
<td>0.4983</td>
<td>0.4927</td>
<td>0.1154</td>
<td>0.187</td>
</tr>
</tbody>
</table>

The top four systems obtained F1 scores between 60% and 69%; the 5th and 6th best system, however, obtained significantly lower F-scores. Our results were lower than the 3 best systems but still above the expect 50% baseline. In terms of recall, our system was also ranked as the 4th best and as to the accuracy results, our system was the 5th best. It obtained performance comparable to the 2nd, 3rd and 4th best systems: 49.14% against 50.00%, 50.06% and 51.71% accuracy. The best system obtained significantly higher accuracy scores compared to all other systems, 64.46% accuracy.

In table 2 we present the best results obtained by the six systems at the error diagnosis level.

<table>
<thead>
<tr>
<th>Team</th>
<th>Accuracy</th>
<th>Precision</th>
<th>Recall</th>
<th>F1</th>
</tr>
</thead>
<tbody>
<tr>
<td>CIRU</td>
<td>0.4589</td>
<td>0.4548</td>
<td>0.4137</td>
<td>0.4333</td>
</tr>
<tr>
<td>NTOU</td>
<td>0.2074</td>
<td>0.2932</td>
<td>0.4149</td>
<td>0.3436</td>
</tr>
</tbody>
</table>
The error diagnosis level is more difficult than the error detection step. This is due to the multiple tags (e.g. missing words, word order) that could be attributed to each instance. At this stage, the performance of all systems was substantially lower than the error detection step. Once again our system was ranked 4th in terms of both F-score and accuracy achieving 23.37% F1 and 25.29% accuracy. The best system achieved 43.33% f-measure and 45.89% accuracy.

The dataset itself was to our understanding very challenging for the frequency-based methods we proposed. We found that some instances were virtually impossible to correctly tag. Examples of instances that were difficult to classify include single words: 老師 (EN 'teacher'), short expressions: 又很貴 (EN 'also very expensive') and instances that without context were difficult to understand even for native speakers: 姓本多 (EN literally: 'nature', 'by itself', 'many')

The results we obtained were consistently ranked in the middle of the table and they are, to our understanding, comparable to the state-of-the-art performance for the task. By looking at the performance obtained by the CIRU team, we see, however, room for improvement, as will be discussed in the next section.

5. Conclusion

This paper described the UDS submission to the shared task on Grammatical Error Diagnosis for Chinese as Foreign Language. We approached the task using frequency information and report results comparable to other state-of-the-art systems. The task is by no means trivial and the almost 9 percentage points behind the best system (CIRU team) showed us that there is still room for improvement. Even so, considering the lack of suitable training data, we believe that the results we obtained are still interesting to report.

We believe that better results can be obtained, for example, by integrating spell checkers (Lin and Chu, 2013) to our algorithm, particularly those that take phonetics into account (Zampieri and de Amorim, 2014). Another issue that should be taken into account in future experiments is the question of segmentation. Very good performance in tokenization is paramount when dealing with logographic languages and this was unfortunately not obtainable with the methods we used.

Finally, in the future we would like to perform experiments to increase the size of the training corpus using artificial errors as proposed by Felice and Yuan (2014). We believe that this is an effective way of producing more data for this task. The performance of these methods when applied to Chinese is still an open question.

Acknowledgements

We would like to thank the shared task organizers, especially Lung-Hao Lee, for providing prompt replies to all our inquiries. We acknowledge that no LaTeX template was available for this submission and we did our best to improve the quality of the formula and tables presented in this paper.

References


Description of NTOU Chinese Grammar Checker in CFL 2014

Chuan-Jie Lin and Shao-Heng Chan
Department of Computer Science and Engineering,
National Taiwan Ocean University, Taiwan, R.O.C
cjlin@ntou.edu.tw

Abstract: This paper describes our first Chinese grammar checker participating in CFL 2014. Several features related to grammatical errors were proposed, including numbers of infrequent word bigrams and POS bigrams. Two SVM classifiers were trained and two formal runs were submitted, where the best F-scores were 66.67% in detection level and 34.36% in identification level.

Keywords: Chinese grammar checking, foreign language learning, machine learning

1. Introduction

Grammar checking for learning Chinese as a foreign language is a new challenge. Mistakes made by foreign students may greatly differ from the ones made by native speakers. It is necessary to study how to build a grammar checker for text written by students who learn Chinese as a foreign language.

As a shared task of ICCE, CFL 2014 (Grammatical Error Diagnosis for Learning Chinese as a Foreign Language) attempts to provide a benchmark to develop techniques on Chinese grammar checking. Four types of errors were defined in this task: redundant word, missing word, word disorder, and word selection problems. In this year, the task only focused on error detection and classification.

- **Redundant word**: a word should be deleted from this sentence
- **Missing word**: a missing word should be added into this sentence
- **Word disorder**: at least one word should change its location in this sentence
- **Word selection**: a word should be replaced into another word

This paper is organized as follows. Section 2 expresses some ideas we have got after observing examples in the training set. Section 3 gives definitions of features. Section 4 delivers experimental results and Section 5 concludes the paper.

2. Observation in Training Data

After observing example sentences in the training set, we found that the occurrence of low-frequency bigrams in the sentence is helpful. We used Google Web IT 5-grams\(^1\) (Google N-grams for short hereafter) as the resource of bigram frequencies. Some examples selected from the training set are provided here to illustrate our hypothesis.

(1) Example of redundant word problem

A redundant word is often unlikely to appear in that context. Moreover, removing this redundant word will create a higher-frequency bigram. For example,

---

\(^1\) https://catalog.ldc.upenn.edu/LDC2006T13
可是 现在*最近 我 工作 很 忙
(But now* recently my work is very busy)

The word “現在” (now) has similar meaning with “最近” (recently), thus it is redundant. As an evidence, the bigram “現在+最近” is not collected in Google N-grams but the frequency of the bigram “可是+最近” is 218250.

(2) Example of missing word problem

Some examples provided in the training set are more like “a missing characters”, not “a missing word”. For example,

聽說 你 準備 開 一個 祝 會
(It is said that you prepare to have a wish* meeting*)

The character “慶” is missing, so the word “慶祝會” (celebration) cannot be correctly identified and is broken into two words “祝” and “會”. As an evidence, the bigram “祝+會” is not collected in Google N-grams.

(3) Example of word disorder problem

Word disorder means that order of the words should be re-arranged into a correct sentence. For example,

你*很 久 以前 找 工作 很 幸*苦
(You very long ago found jobs very lucky* difficult)

In Chinese, a long temporal phrase (“很久以前”, “very long ago” in this example) often appears in front of a complete sentence or, in another word, in front of a subject (“你”, “you” in this example). As an evidence, “以前+找” is not collected in Google N-grams but the frequency of the bigram “你+找” is 305477.

(4) Example of word selection problem

Word selection problem is that at least one word should be replaced with another, more appropriate word. For example,

我 真的 高興 你 找到 一*新的 工作 了
(I really am-happy you found one* new DE job LE)

In Chinese, a long temporal phrase (“很久以前”, “very long ago” in this example) often appears in front of a complete sentence or, in another word, in front of a subject (“你”, “you” in this example). As an evidence, “以前+找” is not collected in Google N-grams but the frequency of the bigram “你+找” is 305477.
When mentioning a countable noun in Chinese, quantifiers (量詞) should be used. For example, to say “a job”, you use “一個工作” (one+GE+job), not “一個工作” (one+job). The character “個” (GE) in this example serves as a quantifier.

However, according to CNS14366, the Segmentation Standard for Chinese Natural Language Processing (中央標準局中文分詞標準, Huang et al., 1997) in Taiwan, a number and a succeeding quantifier are segmented into two words, not grouped as one word. Such example is more like a missing problem rather than a word selection problem to us.

3. Error Detection Features

According to the observations described in Section 2, we defined several features to detect grammar errors as follows.

$f_{bi}$: number of infrequent bigrams appearing in the sentence, where “infrequent bigram” is defined as a bigram NOT collected in Google N-grams. We expect that an erroneous sentence containing more infrequent bigrams. We are also interested to see if the number of infrequent bigrams is related to error types.

$f_{POS}$: number of infrequent POS bigrams appearing in the sentence, where “infrequent POS bigrams” were trained from ASBC, a large POS-tagged corpus. Considering a POS bigram $p_1p_2$, if the probability $P(p_2|p_1)$ is less than 0.01, this bigram is an infrequent POS bigram.

$f_{Nf}$: a Boolean feature denoting the occurrence of a number without a succeeding quantifier, where quantifiers are POS-tagged as Nf.

$f_{stop}$: a Boolean feature denoting the occurrence of a stop POS bigram. We defined a stop list of POS bigrams. POS bigrams in the stop list are:

- $VH + T$, a stative intransitive verb (mostly adjective in English) followed by a particle
- $Cbb + DE$, a correlative conjunction followed by a function word “的”
- $VC + Nd\_DATE$, an active transitive verb followed by a date expression

$f_D$: a Boolean feature denoting the occurrence of a key POS, where key POS includes adverbs (D) and temporal nouns (Nd). Examples of disordered words often fall into these two POS classes.

$f_{bic}$: normalized number of infrequent bigrams, i.e. $f_{bi}$ divided by the length of this sentence.

4. Run Submission and Results

Two runs were submitted to the CFL shared task this year. They were classification results from two different classifiers. System01 uses 5 features, $f_{bi}$, $f_{POS}$, $f_{Nf}$, $f_{stop}$, and $f_D$. System02 also uses 5 features, $f_{bic}$, $f_{POS}$, $f_{Nf}$, $f_{stop}$, and $f_D$. The only difference is the normalization of the first feature. Classifiers were trained by using LIBSVM (Chang and Lin, 2011).

Table 1 shows the performance of these two classifiers on training sets. The performances of the two systems are quite similar. Unfortunately, none of the classifiers can identify any word disorder case. System02 can correctly identify 4 word selection cases thus outperforms System01 a little.

<table>
<thead>
<tr>
<th></th>
<th>System01</th>
<th></th>
<th>System02</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Prec.</td>
<td>Recl.</td>
<td>F-1</td>
<td>Prec.</td>
</tr>
<tr>
<td>Redundant word</td>
<td>43.53</td>
<td>41.73</td>
<td>42.61</td>
<td>41.59</td>
</tr>
<tr>
<td>Missing word</td>
<td>43.71</td>
<td>75.51</td>
<td>55.43</td>
<td>44.71</td>
</tr>
<tr>
<td>Word disorder</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
</tbody>
</table>
Table 2 shows the performance of two formal runs predicted by these two classifiers. The two systems have the same ability to detect errors. In fact, all sentences were predicted as “YES” but only half of them were correct. However, System02 achieved better performance in error-type classification thus outperforms System01 again.

Table 2: Performance of formal runs.

<table>
<thead>
<tr>
<th>Submission</th>
<th>FP Rate</th>
<th>Detection Level</th>
<th>Identification Level</th>
</tr>
</thead>
<tbody>
<tr>
<td>NTOU-Run1</td>
<td>100</td>
<td>50</td>
<td>50</td>
</tr>
<tr>
<td>NTOU-Run2</td>
<td>100</td>
<td>50</td>
<td>50</td>
</tr>
</tbody>
</table>

5. Conclusion

This paper describes our first Chinese grammar checker participating in CFL 2014. Six features related to grammatical errors were proposed, including numbers of infrequent word bigrams and POS bigrams. F-scores of formal runs were 66.67% in detection level and 34.36% in identification level. Normalized features seem outperform original numbers.

Because it was our first attempt to build a Chinese grammar checker, the performance was not satisfied. More studies and more features are needed for building a better system in the future.

References


Harnessing ICT for Educational Development in Emerging Developing Countries within the Asia-Pacific Region

Mas Nida MD KHAMBARIa*, Ahmad Fauzi Mohd AYUBb & Mohammad LUKMANc
Organizers of WICTTEE 2014a,b,c
Faculty of Educational Studies, Universiti Putra Malaysia, Malaysiaa,b
University of Muhammadiyah Prof. Dr. Hamka, Indonesiac
*nidakhambari@gmail.com

ICT trends in the educational landscape of emerging developing countries are growing and will continue to advance as technology plays a huge role in driving the development of economic and social sector (Kozma, 2005). Indeed, the development of these sectors are often used as grounds to justify investments for ICT in education. Reciprocally, globalization has brought center the importance of knowledge creation (Kozma, 2011) with the help of ICT, consequently leveraging it with the aforementioned sectors. Current and new ICT trends are revolutionizing the ways knowledge are delivered and transferred, thus bringing several benefits as well as challenges to educators and learners. Especially in developing countries with diverse communities, the benefits and challenges can come in a multitude of forms. Now that “information and knowledge are the new forms of wealth and are the driving force for development” (Anderson, 2010, p. 10), it is interesting to learn how the citizens of developing countries go about their journey of harnessing ICT into the education system.

In response to the growing research diversity among emerging developing nations within the Asia-Pacific region, the Third International Workshop on ICT Trends in Emerging Economies (WICTTEE 2014) is held in conjunction with the 22nd International Conference on Computers in Education, Nara, Japan. WICTTEE 2014 is organized by the SIG on Development of Information and Communication Technology in the Asia Pacific Neighbourhood—DICTAP. The visions of DICTAP are to:

1. Share ideas and best implementation practices related to government policies and incentives aimed at promoting human resource development, technology transfer, effective e-learning strategies and implementation, software and content development suitable for each member of the Asia-Pacific neighborhood;
2. Coordinate and promote community-based e-learning activities, global sharing and management of information and knowledge. Examples of such communities are the Asia-Pacific Society on Computers in Education (APSCE) and the Association of South East Asian Nations (ASEAN); and
3. Coordinate and promote student and staff exchange among Asia-Pacific neighborhood member nations to promote more effective sharing of knowledge and practices.

The missions of DICTAP are to:
1. Connect researchers from emerging developing countries within the Asia-Pacific region to share scholarly findings and professional insights in ICT development in the field of education;
2. Establish networking opportunities among researchers, reduce the research gap between the researchers from more developed and less developed countries; and
3. Foster, enhance and sustain collaborations among these researchers.
WICTTEE 2014 is the third workshop that we are organizing in the hope to realise the aforementioned visions and missions. The workshop is a continuation of our relentless effort to provide a dynamic platform for practitioners and researchers alike to come together to share their country experiences.

We are extremely pleased that practitioners and scholars with university affiliations from Taiwan, Malaysia, Indonesia, Thailand, and Philippines will be congregating in Nara, Japan to present their research findings and share their views at WICTTEE 2014. A total of eight papers will be presented in a full day workshop.

We would like to take this opportunity to thank all the authors who submitted their papers to WICTTEE 2014. We would like to record our sincerest appreciation to our Program Committee Members who dedicated their time and expertise to the most challenging and demanding task of reviewing the paper submissions. Last but not least, we would like to thank DICTAP’s Advisory Committee Members for their wisdom and guidance in making WICTTEE 2014 a reality.

References


Learning and Affect Trajectories Within Newton’s Playground

Juan Miguel L. ANDRES* & Ma. Mercedes T. RODRIGO*

*Department of Information Systems and Computer Science, Ateneo de Manila University, Philippines
*miglimjanapandres@gmail.com, mrodrigo@ateneo.edu

Abstract: Learning trajectories are typical, predictable sequences of thinking that emerge as students develop understanding of an idea. They have been used principally for research on instructional decision-making, but have also played a significant role in conducting research on learning. Affect refers to experiences of feelings or emotions. The affective states of boredom and confusion, in particular, have been of interest to researchers due to their significant relationships with student learning. In an attempt to account for gains or lack thereof among users of educational software, this study investigates these areas to monitor how students in the Philippines use software or how they are feeling compared to parallel studies conducted in the United States. In particular, this study investigates the relationships between learning trajectories and affect among students in the Philippines using Newton’s Playground, a learning game for physics.

Keywords: Learning trajectories, affect trajectories, Newton’s Playground

1. Theoretical Framework: Learning Trajectories and Affect

Learning trajectories (LTs) represent the “paths by which learning might proceed (Simon, 1995).” They are “typical, predictable sequences of thinking that emerge as students develop understanding of an idea (Daro, Mosher, & Corcoran, 2011).” LTs have been a topic of interest in recent years, enabling researchers to gain a better understanding of student learning. The study of LTs is less than two decades old (Clements & Sarama, 2004), and has only recently been getting attention in the field of learning sciences. Studies show that as researchers and teachers make sense of learning trajectories, they in turn can support growth in knowledge and further student learning.

Studying learning in terms of affective factors has also been a topic of interest in recent years. Two affective states of interest to researchers are confusion and boredom. Confusion or cognitive disequilibrium is the uncertainty about what to do next (D’Mello, Craig, Gholson, Franklin, Picard, & Graesser, 2005). It is interesting because it has a positive and negative dimension (D’Mello & Graesser, 2012), wherein it either spurs learners to exert effort deliberately and purposefully to resolve cognitive conflict, or leads learners to become frustrated or bored, and may lead to disengagement from the learning task altogether (D’Mello & Graesser, 2012).

Boredom, on the other hand, is defined by Fisherl (1993) as an “unpleasant, transient affective state in which the individual feels a pervasive lack of interest in and difficulty concentrating on the current activity.” It has been a topic of interest because of the negative effects usually associated with it, like poorer learning and problematic behaviors, such as gaming the system (Baker, D’Mello, Rodrigo, Graesser, 2010; Rodrigo, Baker, & Nabos, 2010).

The extent to which students learn from educational software is influenced by the effectiveness with which they use software and how they feel while using the software. However, we do not always monitor how students use software or how they are feeling, so we cannot always account for gains or lack of gains. This study is an in-depth examination of software usage and affect and their interactions. Specifically, this study seeks to investigate the relationships between learning trajectories and the affective states of boredom and confusion.
2. Methodology

2.1 Participant Profile

We conducted a study to measure the relationship between a variety of affective and cognitive variables. Data was gathered from 60 eighth grade public school students in Quezon City, Philippines. Students ranged in age from 13 to 16. As of 2011, the school had 1,976 students, predominantly Filipino, and 66 teachers. Of the participants, 31% were male and 69% were female. Participants were asked to rate how frequently they played video games and watched television on a scale of 1 (not at all) to 7 (everyday, for more than 3 hours), and the resulting average frequency of gameplay is 3.2 (in between a few times a month, and a few times a week), and the resulting average frequency of watching television is 5.9 (in between everyday, but for less than 1 hour, and everyday, for 1-3 hours). Participants were asked for their most frequent grade on assignments, and on a scale of 0 (F) to 4 (A), the average most frequent grade of the participants is 3.1 (B).

2.2 Newton’s Playground

Newton’s Playground (NP) is a computer game for physics patterned after Crayon Physics Deluxe. It was designed to help secondary school students understand qualitative physics (Shute, Ventura, & Kim, 2013). Qualitative physics is a nonverbal conceptual understanding of how the physical world operates, along the lines of Newtonian physics. Qualitative physics is characterized by an implicit understanding of Newton’s three laws: balance, mass, and conservation and transfer of momentum, gravity, and potential and kinetic energy (Shute et al., 2013).

NP is a two-dimensional computer-based game that requires the player to guide a green ball to a red balloon by drawing simple machines on the screen with colored markers controlled by the mouse. An example level is shown in Figure 3.1. The player uses the mouse to nudge the ball to the left and right (if the surface is flat), but the primary way to move the ball is by drawing or creating simple machines on the screen with the mouse and colored markers. The objects come to life once the object is drawn. Everything obeys the basic rules of physics relating to gravity and Newton’s three laws of motion (Shute et al., 2013).

![Figure 1. Example level of Newton’s Playground.](image)

The 74 levels in NP require the player to solve the problems via drawing different simple machines, representing agents of force and motion: inclined plane/ramps, levers, pendulums, and springboards. Again, all solutions are drawn with colored markers using the mouse. A ramp is any line drawn that helps to guide a ball in motion. A ramp is useful when a ball must travel over a hole. A lever rotates around a fixed point, usually called a fulcrum or pivot point. Levers are useful when a player wants to move the ball vertically. A swinging pendulum directs an impulse tangent to its direction of motion. The pendulum is useful when the player wants to exert a horizontal
force. A springboard (or diving board) stores elastic potential energy provided by a falling weight. Springboards are useful when the player wants to move the ball vertically.

**Gold badges versus silver badges.** Some levels in NP have multiple solutions, which means a player can solve the level using different agents. Gold badges are awarded when a player solves a problem “under par”, that is, under a limit set for a specific solution. For example, a level may be solved using a ramp, with a par of 1 object, or a pendulum, with a par of 3 objects. If a player solves the level with more objects than par, he receives a silver badge. Gold badges suggest that the player has mastered the agent relevant to the given level. Silver means the player may not have fully mastered the agent yet.

2.3 **NP Interaction Logs**

We collected two types of data during the study: interaction logs and human observations. During gameplay, NP automatically generates interaction log files. Each level a student plays creates a corresponding log file, which tracks every event that occurs as the student interacts with the game. The events per level and their respective attributes that are relevant to this study are:

- Level Start,
- Level Restart,
- Level End, an event that signals the player solved the level,
  - Badge, an attribute that states the type of badge (i.e. gold or silver) awarded to the player
  - Agent, an attribute that states for which agent the badge was awarded for
- Menu Focus, an event that signals the player gave up and quit the level without solving it,
- Drawing of any of the four agents,
- Object Limit, an event that is triggered by the player reaching the maximum number of objects drawn, and
- Stacking, an event that signals the player is gaming the system.

Each of these features provides useful information about students’ gameplay behaviors, which can then be used to make inferences about how well they are doing in the game (Shute et al., 2013).

2.4 **The Baker-Rodrigo-Ocumpaugh Monitoring Protocol**

The Baker-Rodrigo-Ocumpaugh Monitoring Protocol (BROMP) is a protocol for quantitative field observations of student affect and behavior. BROMP is a holistic coding procedure that has been used in thousands of hours of field observations of students, from kindergarten to undergraduate populations. It has been used for several purposes, including to study the engagement of students participating in a range of classroom activities (both activities involving technology and more traditional classroom activities) and to obtain data for use in developing automated models of student engagement with Educational Data Mining (EDM) (Ocumpaugh, Baker, & Rodrigo, 2012). Within BROMP, each student observation lasts 20 seconds, and the observers move from one student to the next in a round robin manner during the observation period.

The affective states observed within Newton’s Playground were concentration, confusion, frustration, boredom, happiness, delight, and curiosity. The behaviors observed were on-task, off-task, stacking, and a behavior called without thinking fastidiously (WTF), a behavior in which, despite a student’s interaction with the software, “their actions appear to have no relationship to the intended learning task (Wixon, Baker, Gobert, Ocumpaugh, & Bachmann, 2012).” The inter-coder reliability for affect was acceptably high with a Cohen’s (1960) Kappa of 0.67. The typical threshold for certifying a coder in the use of BROMP is 0.6, established across dozens of studies as well as the previous affective computing literature.
2.5 The Human Affect Recording Tool

The Human Affect Recording Tool, or HART, is an Android application developed to guide researchers in conducting quantitative field observations according to the BROMP protocol. The application synchronizes the coded observations to Internet time, allowing for more precise synchronization with log file data from the educational software under study.

HART asks for input regarding school and classroom information, coding schemes to be used, and the student IDs of the students to be observed during the session. The application then presents the student IDs in the order entered, allowing BROMP observers to more conveniently code affect and behavior until the session is manually terminated. All observations are logged on a text file that is locally stored on the device used to run HART. The application and all its functions are discussed in more detail in (Ocumpaugh et al., 2012).

2.6 Data Gathering Process

Before playing NP, students completed a 16-item multiple-choice pretest for 20 minutes. Students were then assigned a computer on which they would play NP. Students played the game for two hours, during which, two trained observers used BROMP to code student affect and behavior. A total of 36 observations per participant per observer were collected. Videos of participants’ faces were also recorded during gameplay. After completing the two hours of gameplay, participants completed a 16-point multiple-choice posttest for 20 minutes. The pretest and posttest were designed to assess knowledge of physics concepts, and has been used in previous studies involving NP (Shute et al., 2013).

In order to investigate learning within Newton’s Playground, we made use of the interaction logs recorded during gameplay to analyze student performance. Of the 60 participants, data from 12 students were lost because of faulty data capture and corrupted log files. Only 48 students had complete observations and logs. The analysis that follows is limited to these students.

The BROMP observations were tabulated, and the percentage of each affective state per student was calculated. Boredom, confusion, and frustration were three of the more commonly observed affective states, besides concentration.

All interaction logs were passed through a parser to arrange log events neatly in tab-delimited text files. These text files were then run through a filter to get per student, per level, per attempt summaries, such as total time spent, total number of restarts, total number of objects drawn, etc. Finally, the information was collapsed to form per student vectors that summarized the students’ entire interactions with the game. All statistical analyses conducted within this study are limited to the computation of percentages and result visualization.

3. Findings

We collected pre-test and post-test data from each student (N=60). Scores were generally poor. Students averaged 6.02 correct answers on both the pre-test and the post-test, out of a highest possible score of 16. This indicates that no learning improvements were detected. What follows is a descriptive analysis of the gathered data using methods described in the following subsections.

We operationalize learning trajectories on two levels:

1. On a coarse-grained level, learning trajectories are the performances of students in terms of gold and silver badges earned during gameplay, and
2. On a fine-grained level, learning trajectories are the sequences of students’ interaction behaviors while solving or not solving a level.

As mentioned previously, affect coders followed BROMP, which resulted in 36 observations per student, per observer. For the purposes of this study, we define the incidence of affect as the percentage of students observed to be in a specific affective state during one observation count. We operationalize affect trajectories as the incidence of affect over time, that is, over the span of the 36 observations.
The findings in this section are from analyses conducted in finding:
1. The players’ coarse-grained learning trajectories within NP,
2. The players’ boredom and confusion trajectories, and
3. The relationships mined between the two.

3.1 Coarse-Grained Learning Trajectory Analysis

For the coarse-grained LT analysis, the percentages of students earning gold, silver, or no trophies were graphed over their opportunities to practice each of the four agents.

The three performance metrics (i.e. earning a gold trophy, earning a silver trophy, and earning no trophy) were used to track how well a student performed during gameplay, and in turn, see how well they understood each of the four agents used in the game. Every time a badge is awarded to a student, it is awarded for a specific agent. If a ramp was used to earn a gold badge, the student will get a gold badge for the ramp for that level. This is especially important for levels wherein any of the agents can be used to solve a level. Most levels, as the data showed however, only award badges for one of the four agents.

Using the logs generated by NP, trophies were grouped by level and by agent. In doing so, we were able to track which agents were awarded medals per levels, thus determining which agents were needed to solve each of the levels. Table 1 shows the tally of the first ten levels.

Table 1: Tally of trophies, by level and by agent for the first ten levels.

<table>
<thead>
<tr>
<th>Level</th>
<th>Ramp</th>
<th>Lever</th>
<th>Pendulum</th>
<th>Springboard</th>
</tr>
</thead>
<tbody>
<tr>
<td>P01L01</td>
<td>60</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P01L02</td>
<td>47</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P01L03</td>
<td></td>
<td>51</td>
<td></td>
<td></td>
</tr>
<tr>
<td>P01L04</td>
<td></td>
<td></td>
<td>48</td>
<td></td>
</tr>
<tr>
<td>P01L05</td>
<td></td>
<td></td>
<td></td>
<td>51</td>
</tr>
<tr>
<td>P02L01</td>
<td>47</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P02L02</td>
<td></td>
<td></td>
<td>46</td>
<td></td>
</tr>
<tr>
<td>P02L03</td>
<td>46</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P02L04</td>
<td></td>
<td>22</td>
<td>46</td>
<td>30</td>
</tr>
<tr>
<td>P02L05</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

This table shows which agents levels awarded badges for, whether gold or silver, which in turn gives us an idea of what agents were intended to solve the levels. The solution to Playground 1, Level 1, for example, is a ramp, with all sixty attempts on the level being awarded with ramp badges.

Using this table, we selected the first ten levels in which each of the four agents was used as a solution. Each of these first ten levels was then treated as an opportunity to practice one of the four agents. An opportunity to practice refers to a chance given to the student to exercise a specific skill, e.g. constructing a pendulum that pushes the ball to the target. Every level is solvable using one or more of these agents; therefore every level has one or more opportunities to practice possibly a variety of skills.
It is important to note that students were free to choose the levels they wanted to solve. The software did not force them through the material in a stepwise fashion. Furthermore, levels were not grouped thematically, by agent, so even if a student solved the levels sequentially, he would have opportunities to practice different agents.

The percentage of students that earned gold trophies, silver trophies, and no trophies for each of the opportunities was then graphed. Figures 2 and 3 show the students’ learning trajectories for each of the four agents. However, these graphs need to be unpacked further because students could choose what levels to solve, the actual levels corresponding to opportunities to practice 1 through 10 varied per student. That is for example, student 1’s first opportunity to practice may be different from student 2’s.

A consistent pattern can be observed across all four graphs, that is, as students progress through each of their ten opportunities to practice each of the four agents within NP, the perform more poorly over time. The percentage of students earning gold and silver trophies decreases over time, while the percentage of students unable to solve levels increases.
3.2 Fine-Grained Learning Trajectories

For the fine-grained LT analysis, a sequence mining analysis is to be conducted, taking into consideration the common sequences of actions student took in trying to solve each level.

As previously mentioned, NP generated interaction log files per level attempt that track every event that occurs as the player tries to solve the level. A filter was developed to pull only the relevant events from the log files. This filter is to be run on each level, arranging events chronologically on an output text file, divided by student. Events are to then be placed on a previous state-current state table to track transitions and transition frequencies between states. Using frequency calculations, common paths can easily be graphed and tracked through interaction network diagrams. This analysis is still in progress.

3.3 Affect Trajectory Analysis

As seen in the results above, students performed more poorly as they progressed through the levels in the game, and that at a certain point, the number of students earning trophies would just continuously decrease. The hypothesis this analysis sought to prove was that affect experienced by the students during gameplay could somehow be related to the students’ eventual poor performance.

Using the logs generated by HART, all human observations per student were lined up on an Excel sheet. Each observer had a total of 36 observations per student. Using both of the observers’ logs, a total of 72 observations per student were recorded. The percentage of students who were observed to be bored and the percentage of students who were observed to be confused per observation were calculated. An average percentage between the two coders was then calculated for each of the 36 observations. Figure 4 shows the affect trajectories of both confusion and boredom over time.

It is interesting to note that while confusion was experienced by a steady number of students during the entire 2-hours of gameplay, the number of students experiencing boredom increased as the session progressed. The increase in percentage of boredom begins at observation 21, which is about one hour and ten minutes into the session.

3.4 Relationships

This study hopes to examine two relationships:
1. Between the coarse-grained LTs and affect, finding relationships between overall in-game student performance and incidences of both boredom and confusion, and
2. Between the fine-grained LTs and affect, finding relationships between common sequences and incidences of both boredom and confusion.

3.4.1 Coarse-Grained LTs and Affect

The BROMP observations were tabulated, and the percentage of each affective state per student was calculated. All interaction logs were passed through a parser to arrange log events in tab-delimited text files. These text files were then run through a filter to get per student, per level, per attempt gameplay features, such as total time spent, total number of restarts, total number of objects drawn, etc. Finally, the information was collapsed to form per-student vectors that summarized the students’ entire interactions with the game. Each vector included the following performance metrics:

- **Gold badge** – percentage of level attempts solved, earning the student a gold badge
- **Silver badge** – percentage of level attempts solved, earning the student a silver badge

These two metrics were correlated with the students’ respective percentages of boredom. The analysis, however, found no significant relationships. A previous study that ran the same methodology, however, found significant correlations between these metrics and confusion (Andres et al., in press). The study reported confusion to be negatively correlated with earning a gold badge, but positively correlated with earning a silver badge.

3.4.2 Fine-Grained LTs and Affect

This sequence mining analysis will take into consideration the common sequences mined in the previous analysis (in 3.2), and correlate them with the percentages of time the students were observed to be either confused or bored. We hypothesize that some sequences will be characteristic of either affective state, and can then be used as indicators within the game. As with the analysis in 3.2, this analysis is still in progress.

4. Discussion, Conclusions, and Future Work

The study attempted to identify learning and affect trajectories among students using an educational game for Physics, called Newton’s Playground. In each level of the game, players are made to get a green ball to a red balloon using one or a combination of these four simple machines: lever, ramp, springboard, and pendulum.

The study operationalized learning trajectories (LTs) on two levels: coarse-grained LTs track the students’ performance in terms of gold and silver badges earned, and fine-grained LTs track in-game events that occur as students interact with the game. Four coarse-grained learning trajectories were analyzed, one for each of the four simple machines in the game. All four coarse-grained LTs showed a common pattern of students performing more poorly as time progressed, earning less badges, and solving less levels. The fine-grained LT analysis is still in progress.

The study also looked at boredom and confusion trajectories among the students. Results showed that while confusion was experienced by a steady amount of students throughout the 2 hours of gameplay, the percentage of students experiencing boredom increased over time.

The study attempted to find relationships between the learning trajectories and affect, and in doing so, found no significant relationships between performance and boredom. A previous study found significant correlations with confusion, however, where confusion was negatively correlated to earning a gold badge, and positively correlated with earning a silver badge (Andres et al., in press). The analysis between fine-grained LTs and affect is still in progress.

We speculate that there are a number of relationships that are worth further exploration. ICT has not successfully penetrated the education system in the Philippines. Several infrastructural, financial, and implementation hindrances still exist, and despite the government’s best efforts to work around them, programs and projects still fall through the cracks. Several
government projects are currently in place, however, that aim to 1) ease the integration of ICT in the classroom for both teachers and students, 2) help alleviate poverty, most of which harness technology to maximize outcomes, and 3) utilize technology to reach potential learners who don’t have immediate access to any form of formal learning.

On the student level, poor prior knowledge (as evidenced by students’ poor pre-test results) might have made the game daunting, causing the students’ poor performance in the game over time. The game interaction time of two hours may have been too long, leading to the increase in boredom. Indeed, the researchers noticed that the students rushed through the post-test, implying that they wanted to leave the testing area as quickly as possible. Boredom might have led to systematic guessing and other similar non-learning behaviors, leading in turn to poor post-test scores (Baker et al, 2010). In future work, we intend to verify which among these hypotheses the data support. In doing so, we hope to contribute to principles that guide the development of good educational games.

Acknowledgements

We would like to thank the Ateneo Center for Educational Development, Carmela C. Oracion, Christopher Ryan Adalem, the officials at Krus Na Ligas High School for making this study possible. We would also like to thank Ryan Baker, Valerie Shute, Luc Paquette, Matthew Ventura, Matthew Small, and the Gates Foundation Grant #OP106038 for collaborating with us.

References

Assessing Organizational Support and System Characteristics of Learning Management System: Views from Malaysian Higher Education Undergraduate Student

*Sousan BALEGHI ZADEH, Ahmad Fauzi MOHD AYUB, Rosnaini MAHMUD, Shaffee MOHD DAUD

*Faculty of Educational Studies, Universiti Putra Malaysia
bcInstitute of Mathematical Research, Universiti Putra Malaysia

s_baleghi@yahoo.com, afmy@upm.edu.my, rosnaini@upm.edu.my; shaffee@upm.edu.my

Abstract: Learning management system (LMS) is a kind of software that supports teaching and learning activities. Today, the number of educational institutions equipped with LMS is increasing. However, there are many factors that make students reject or accept this kind of technology. The present study assessed two characteristics of LMS (organizational support and system characteristics) in views of students. These two characteristics were assessed by three variables (technical support, system interactivity and system functionality). The respondents of the present study were 216 undergraduate students of faculty of education in Universiti Putra Malaysia. The result of the study revealed that in views of the respondents, organizational support was moderate, while system characteristics were high.

Keywords: Learning Management System, System Characteristics, Organizational Support

1. Introduction

One of the popular concepts that ICT has produced in the realm of education is e-learning (Hernandez et al., 2011; Šumak, Heričko, & Pušnik, 2011). There are many definitions for e-learning. Some of them are broader and encompass different types of ICT, while the others are narrower. For example, Hill and Wouters (2010) have defined e-learning as "use of ICTs (e.g. Internet, Intranet, CD-Rom, interactive TV, teleconferencing, computer conferencing and chat) to deliver instruction to learner" (p.204). However, some definitions of e-learning are narrower. For example, according to O’Mahony (2004) and Chang (2008) e-learning refers to any form of instruction delivered just through the web. Systems that conduct e-learning are different and have various names, such as online systems, virtual systems, learning management systems and so on (Piotrowski, 2010). To avoid getting confused, in the present study the term learning management system (LMS) is used for any kind of e-learning systems. LMS is one of the most popular software in that its usefulness in higher education institutions is widely increased (Chang, 2008; Dutta, Roy, & Seetharaman, 2013). Learning management system is a kind of information system that supports teaching and learning (Dutta et al., 2013). In fact, it organizes and provides tools through which students will be able to download learning contents, build, and deliver online learning environments (Piña, 2012). One of the most important benefits of LMS is to generate and manage reports on learners and assessment results (Theis, 2005). Besides, through the features of LMS, instructors and students can convey instructional materials, send notice to class, submit assignments, and interact with students (Lonn & Teasley, 2009). In fact, this information system combines technology features and pedagogy (Ioannou & Hannafin, 2008).

Although investing on LMS in institutional education is enhancing, research has reported that faculty and teachers are not interested in using technology (Chang, 2008; Hadjipavli, 2011). There are many factors which may affect LMS utilization of students and lecturers. However, Davis,
Bagozzi, and Warshaw (1989) suggest that organizational characteristics and system characteristics are the crucial factors that may influence information system utilization. Technical support which is sometimes called facilitating support (Venkatesh & Bala, 2008) belongs to organizational characteristics. Technical support assists users to solve problems they encounter when they are working with an information system (Ngai et al., 2007). This factor enhances user satisfaction and has a critical influence on beliefs of users in accepting or rejecting an information system (Igbaria, Guimaraes & Davis, 1995). There are several studies in the LMS environment which indicate that technical support had a significant effect on LMS utilization. For example, Ngai et al. (2007) investigated the acceptance of LMS among undergraduate and postgraduate students of seven universities in Hong Kong with a sample size of 1263 and found out that technical support had a significant effect on LMS utilization. In another study, Sánchez and Hueros (2010) also examined LMS acceptance (Moodle) among students of business management and educational sciences in the University of Huelva (Spain) with a sample size of 226. The results of this research revealed that technical support had a significant effect on system utilization.

System characteristics encompass different variables such as system functionality and system interactivity (Davis et al., 1989; Pituch & Lee, 2006). Indeed, system functionality assesses the flexibility and quality of LMS features from the users’ point of view, e.g. whether LMS is equipped with features through which students can send their assignments, download the contents of syllabus, take quizzes, use a variety of media such as audio, text and video and so on (Pituch & Lee, 2006; Selim, 2003). System interactivity refers to interaction among instructors and students in the process of learning and teaching (Pituch & Lee, 2006). The LMS should be equipped with features such as forum, email and chat room, through which students and teachers can interact with each other. Indeed, this factor assesses the interaction between lecturers and students (Pituch & Lee, 2006). There are several studies which indicate that system functionality and system interactivity had a significant effect on system usage. For example, Pituch and Lee (2006) investigated the influence of system characteristics (system interactivity and system functionality) on LMS utilization among 251 Taiwanese college students and found out that both system interactivity and system functionality had a significant effect on LMS utilization. Moreover, these variables obtained a high mean value. Wang and Wang (2009) also investigated the influence of system characteristics among 268 university instructors of Taiwan and revealed that system characteristics had a significant effect on LMS utilization. The outcomes of this research also indicated that system characteristics obtained a high mean value. The main purpose of the present study is to assess organizational characteristics and system characteristics of learning management system of Universiti Putra Malaysia (PutraLMS) in views of undergraduate students.

2. Research Objectives

i. To assess technical support of PutraLMS in views of students.
ii. To assess system interactivity of PutraLMS in views of students.
iii. To assess system functionality of PutraLMS in views of students.

3. Research Methodology

The population of the present study was full-time undergraduate students of faculty of educational studies of Universiti Putra Malaysia (UPM) in the second semester of the academic year 2012-2013. The students were selected through cluster sampling with a sample size of 216. The design of the present research is also descriptive.

4. Research Instrument

The variables of the present study were technical support, system interactivity, and system functionality and they were measured through a questionnaire with 23 items. Technical support is intended to measure services assisting undergraduate students of education at UPM to solve hardware and software problems with PutraLMS. The six items used to measure this construct were adopted from Sánchez and Huerous (2010), and Ngai et al. (2007). The second construct is
system interactivity. This construct is used to measure the ability of PutraLMS in providing facilities for interaction. This entails interaction among undergraduate students themselves, interaction between lecturers and students, and collaboration in learning which results from these interactions. This construct includes seven items adopted from Pituch and Lee (2006) and also self-developed items. It should be noted that self-developed items refer to items which were created by authors according to the definition of the variables. The last construct is system functionality which assesses undergraduate students’ perception of flexibility of PutraLMS (UPM) in accessing instructional and media. This construct consists of 10 items adopted from Selim Ahmed (2010), Pituch and Lee (2006), and self-developed items. All items for this construct were measured through 5-point Likert-scale items labelled as 1 (strongly disagree), 2 (disagree), 3 (not sure), 4 (agree) and 5 (strongly agree). Two experts of the faculty of education at UPM examined the content validity of the questionnaire and their comments were followed. Reliability of the questionnaire was measured by Cronbach’s alpha. Reliability of each variable was: technical support (.90), system interactivity (.91) and system functionality (.82). Since some items of the questionnaire were self-developed, its content validity was examined by four experts of education at Universiti Putra Malaysia.

5.0 Research findings

The results of the present study are based on descriptive study. The variables of the study are technical support, system interactivity, and system functionality.

5.1 Overall Mean

Table 1 indicates the overall mean of the variables of the study (technical support, system interactivity and system functionality). Among the three variables, system functionality obtained the highest mean (Mean = 3.83, SD = .53). This is followed by system interactivity (Mean = 3.53, SD = .74) and technical support (Mean = 3.35, SD = .67). The next section will discuss in detail items used to measure all the variables in this study.

Table 1. Overall Mean of Variables

<table>
<thead>
<tr>
<th>Factors</th>
<th>Number of items</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Technical Support</td>
<td>6</td>
<td>3.35</td>
<td>.67</td>
</tr>
<tr>
<td>System Interactivity</td>
<td>7</td>
<td>3.53</td>
<td>.74</td>
</tr>
<tr>
<td>System Functionality</td>
<td>10</td>
<td>3.83</td>
<td>.53</td>
</tr>
</tbody>
</table>

S.D.: standard deviation

5.2 Technical Support

When using any information system, technical support refers to assist users to solve problems they encounter when they are working with an information system (Ngai et al., 2007). Table 2 indicates the mean and standard deviation of six items of technical support whereby the highest mean refers to PutraLMS offers good technical support (Mean = 3.42, S.D. = .820) followed by the manual on the operation of PutraLMS is sufficient (Mean = 3.38, S.D. = .887). There are two items which shared the same mean, which refers to e-mail inquiries to the technical support group when facing technical problem while using PutraLMS (Mean =3.35, SD = .799) and there is a hotline for fixing user problems (Mean = 3.35, S.D. = .787). Overall, the respondents felt that technical support is sufficient for them to use PutraLMS during their study period.

Table 2: Technical Support

<table>
<thead>
<tr>
<th>Items</th>
<th>Source</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>A hotline for fixing user problems is available at any time in PutraLMS.</td>
<td>Sánchez &amp; Huerous, 2010</td>
<td>3.35</td>
<td>.787</td>
</tr>
<tr>
<td>I can rely on the technical support group while using PutraLMS.</td>
<td>Self-developed</td>
<td>3.28</td>
<td>.823</td>
</tr>
</tbody>
</table>
Email inquiries to the technical support group can be made when there is a technical problem while using PutraLMS.  
Sánches & Huerous, 2010  3.35  .799

Web-based inquiries can be made when there is a technical problem while using PutraLMS.  
Sánches & Huerous, 2010  3.34  .853

The manual on the operation of PutraLMS is sufficient.  
Ngai, Poon & Chan, 2007  3.38  .887

PutraLMS offers good technical support.  
Sánches & Huerous, 2010  3.42  .820

TS: technical support; S.D: standard deviation

5.3 System Interactivity

Interactivity is an important aspect when users interact with information system. Therefore, system interactivity is the ability of the system to provide opportunities for interaction among users (Pituch & Lee, 2006). This section will describe students’ perception towards the ability of PutraLMS in providing facilities for interacting among students, the interactions between lecturers and students, and collaboration in learning which grows out of these interactions. Analysis towards seven items shows that the highest mean refers to students’ perception towards PutraLMS enabling interactive communication between lecturers and students (Mean = 3.65, S.D. = .93) followed by PutraLMS enabling them to receive comments (Mean = 3.61, S.D. = .86) and features of collaborative learning in PutraLMS (Mean = 3.60, S.D. = .90) (refer to Table 3). Overall, the students felt that PutraLMS provide features that enable them to interact among the colleagues and lecturers.

<table>
<thead>
<tr>
<th>Item</th>
<th>Source</th>
<th>Mean</th>
<th>S.D.</th>
</tr>
</thead>
<tbody>
<tr>
<td>PutraLMS enables interactive communication between lecturers and students.</td>
<td>Pituch &amp; Lee, 2006</td>
<td>3.65</td>
<td>.93</td>
</tr>
<tr>
<td>I can see the features of collaborative learning (e.g. group work) in PutraLMS.</td>
<td>Self-developed</td>
<td>3.60</td>
<td>.90</td>
</tr>
<tr>
<td>The communication tools (email, forum, chatroom, etc.) in PutraLMS are effective.</td>
<td>Pituch &amp; Lee, 2006</td>
<td>3.51</td>
<td>.90</td>
</tr>
<tr>
<td>PutraLMS enables me to receive my lecturers’ comments.</td>
<td>Self-developed</td>
<td>3.61</td>
<td>.86</td>
</tr>
<tr>
<td>PutraLMS enables interactive communication among students.</td>
<td>Pituch &amp; Lee, 2006</td>
<td>3.49</td>
<td>.90</td>
</tr>
<tr>
<td>I can share my knowledge with my classmates through PutraLMS.</td>
<td>Pituch &amp; Lee, 2006</td>
<td>3.53</td>
<td>.87</td>
</tr>
<tr>
<td>My lecturers often communicate with us through PutraLMS.</td>
<td>Pituch &amp; Lee, 2006</td>
<td>3.37</td>
<td>1.07</td>
</tr>
</tbody>
</table>

SI: system interactivity; S.D.: standard deviation

5.3 System Functionality

System functionality is flexibility of an information system (Pituch & Lee, 2006). Therefore, this section will investigate students’ perception of flexibility of PutraLMS in accessing instructional and assessing media. The construct of system functionality was measured by 10 items. The highest mean refers to the capability of PutraLMS to print course materials (Mean = 4.33, S.D. = .714). Students also felt that they can access PutraLMS from any place (Mean = 4.14, S.D. = .851) and it offers flexibility in learning regarding time (Mean = 4.08, S.D. = .742)

<table>
<thead>
<tr>
<th>Items</th>
<th>Source</th>
<th>Mean</th>
<th>S.D.</th>
</tr>
</thead>
<tbody>
<tr>
<td>I can print course materials through PutraLMS.</td>
<td>Selim, 2010</td>
<td>4.33</td>
<td>.714</td>
</tr>
<tr>
<td>PutraLMS offers flexibility in learning regarding time.</td>
<td>Pituch &amp; Lee, 2006</td>
<td>4.08</td>
<td>.742</td>
</tr>
<tr>
<td>I can access PutraLMS from any place.</td>
<td>Pituch &amp; Lee, 2006</td>
<td>4.14</td>
<td>.851</td>
</tr>
</tbody>
</table>
PutraLMS offers different types of material (e.g., audio, video, and text) for every kind of course content. Pituch & Lee, 2006 3.62 1.019

PutraLMS provides opportunity for taking tests. Pituch & Lee, 2006 3.27 .957

PutraLMS presents course material in a well-organized manner. Pituch & Lee, 2006 3.70 .958

PutraLMS clearly presents course contents. Self-developed 3.93 .804

PutraLMS facilitates groupwork. Self-developed 3.50 .857

PutraLMS provides opportunity for sending assignments. Pituch & Lee, 2006 3.72 .893

The course material in PutraLMS is in a readable format. Pituch & Lee, 2006 4.05 .649

SF: system functionality; S.D.: standard deviation

5. Discussion

Learning management system sometimes called e-learning platform, e-learning, online learning, and virtual learning emerged when instruction via network was conducted (Chang, 2008; Piotrowski, 2010). LMS provides a variety of opportunities for instructors and learners to increase their educational experiences (Holmes & Gardner, 2006). Today, LMS is widely used in higher education (Dutta, 2013). Nevertheless, the outcomes of many studies indicate that the quality of using LMS by students and faculties is limited (The Campus Computing Survey, 2008). For example, Lam, Lo, Lee, and McNaught (2012) investigated using LMS (WebCT) by undergraduate and graduate students in the Chinese University of Hong Kong and found that only 14.8 percent of the students used features for online discussion. According to Almarashdeh, Sahari, Mat Zin, and Alsmadi (2010), in almost all colleges of Malaysia, many lecturers use LMS just for transferring materials and never use communication features such as discussion board, wiki, chat room and so on. In another research, Embi, Hamat, and Sulaiman (2012) examined LMS utilization among 26 Malaysian university lecturers. The results of their study revealed that only two-thirds of lecturers used LMS, such that 65 percent of utilization was restricted to course delivery.

There are many factors affecting utilization of information system. However, system characteristics and organizational support can be considered as crucial factors that may affect system utilization (Davis et al., 1989; Venkatesh & Bala, 2008). In the present study, the mean value of three variables of technical support, system interactivity, and system functionality were measured. Among these variables, system interactivity and system functionality belong to system characteristics and technical support belongs to organizational system. In the present study, technical support obtained the lowest mean. This suggests that organizational support should increase assistance of students when they encounter hardware or software problems. Although system functionality and system interactivity obtained higher mean, it is suggested that system flexibility and system interactivity of PutraLMS improve.

This study has its own limitations. First the respondents of the study were full time students of faculty of education; therefore, the generalization of the findings should be done by care, because the backgrounds of part time students as well as students from other faculties may be different from full time students. Second, there are many variables such as perceived usefulness, perceived ease of use and subjective norm which may affect LMS utilization of users, but the present study was limited to investigating three variables (technical support, system interactivity and system functionality).

References


Exploring Deep Approach to Learning for Accounting through ICT-Supported Learning Environment in Malaysian Secondary Schools: A Preliminary Study

Boon See TAN\textsuperscript{a} & Su Luan WONG\textsuperscript{b}

\textsuperscript{a}Faculty of Educational Studies, Universiti Putra Malaysia, Malaysia
\textsuperscript{b}Faculty of Educational Studies, Universiti Putra Malaysia, Malaysia
*bstan_29@yahoo.com

Abstract: Under the commitment of the Malaysian Ministry of Education to transform the curriculum and assessment of Principles of Accounting in secondary schools where deep approach to learning is critical for attaining quality learning outcomes, this preliminary study aims to assess the instrument used to measure students’ approaches to learning for Principles of Accounting and perceptions of ICT-supported learning environment. It also aims to explore the current status of students in terms of these two variables, and their relationship. A total of 33 secondary school students who are studying Principles of Accounting in an ICT-supported learning environment participated in this study. They responded to the questionnaire which consists of the subscales of Deep Approach, Surface Approach that measure students’ approaches to learning, as well as subscales of Student Cohesiveness, Teacher Support, Cooperation, Involvement, Investigation, Task Orientation, Computer Usage, and Appropriate Assessment that assess for perceptions of ICT-supported learning environment. The findings indicate that the measures for Surface Approach need to be refined as it only obtained a satisfactory level of reliability value compared to other subscales. The descriptive statistics findings reveal that students exhibited higher deep approach than surface approach to learning. They also showed a rather low score in Appropriate Assessment which may reflect a more fact-and memory oriented assessment was perceived. Further Pearson correlation analysis suggests that Teacher Support and Computer Usage are the main variables to be significantly related to Deep Approach. They are simultaneously associated with other subscales of perceptions such as Student Cohesiveness, Cooperation, Involvement, Investigation, and Task Orientation that are also significantly related to Deep Approach. In addition, most of the subscales of perceptions are related to Deep Approach except Appropriate Assessment. This suggests that the perceived fact and memory-oriented assessment was not significantly associated with Deep Approach.

Keywords: Deep approach to learning, surface approach to learning, perceptions of ICT-supported learning environment, Principles of Accounting

1. Introduction

Accounting is widely referred to as the “language of business”. Its functions are identifying, measuring, and communicating economic information to permit informed judgments and decisions by users of the information (Martin, 1994). The fundamental of accounting encompasses the basic knowledge of accounting in identifying and measuring financial information through the application of the double-entry book-keeping system. In Malaysia, the education of fundamental accounting starts at the upper secondary school level (i.e. Form 4 and 5 or Grade 10 and 11) where the accounting knowledge and skills are delivered through the subject of Principles of Accounting. This subject consists of the concepts, principles, and accounting methods complemented by the
skills in recording, classifying, interpreting, and summarising financial data based on business transactions (Technical Education Department, 2000).

It was observed by few researchers that students always perceive that learning accounting is simply about learning a set of rules and evidences suggest that they tend to adopt a surface learning approach compared to other subjects (Eley, 1992; Beattie, Collins, & McInnes, 1997; Booth, Luckett, & Mladenovic, 1999; Lucas, 2001). Similarly, in the Malaysian context, students’ learning for the subject of Principles of Accounting has yet to achieve deep approach as it was found that most of the accounting teachers tended to use the teacher-centred teaching methods such as lecture, drill and practice, and demonstration of problem solving without delving deeper into the knowledge (Suhaida Abdul Kadir, 2002; Hanuni Yusuf, 2003; Rohaila Yusof, 2006). Such methods could lead to surface learning where the lower-level procedural skills are acquired without processing information for meaning. This may be a reflection of the exam-oriented education system which is the common learning issue faced by Asian countries. It affects many teachers unwilling to take the risk of students’ failure in examination by attempting innovative teaching. They would rather employ the teaching methods which comply with the requirements of the examination system which is mostly teacher-centred (Looi, Hung, Chen, & Wong, 2006). Moreover, students are short of the ICT skills to adopt deep approach to learning for managing the whole set of accounts by relating the processes of accounting cycle into a coherent whole (Arfah Salleh, 2001; Rashidah Hassan & Arfah Salleh, 2008; Tan & Wong, 2012).

In view of the aforementioned weaknesses in the Principles of Accounting education, the Malaysian Ministry of Education has strived to transform its curriculum and assessment from content-based to skilled-based from 2010 onwards where deep approach to learning is critical for attaining the desired learning outcomes such as the master of ICT skills and soft skills (e.g. higher order thinking skills, communication skills, problem solving, and decision making skills) (Bahagian Pembangunan Kurikulum [Curriculum Development Section], 2009). However, to what extent the revised curriculum and assessment are able to foster students’ deep approaches to learning, particularly under the ICT-supported learning context? Thus, a preliminary study was conducted with the following objectives:

• To assess the appropriateness and reliability of the instrument;
• To take the first step in exploring students’ approaches to learning and perceptions of ICT-supported learning environment; and
• To conduct initial investigation on the relationship between students’ perceptions of ICT-supported learning environment and approaches to learning.

2. Literature Review

2.1 Students’ Approaches to Learning (SAL)

SAL is defined as the ways in which a student perceived a particular academic task and then handle it (Marton & Säljö, 1976). In addition, SAL is seen as a contextually dependent response rather than an enduring characteristic of the individual (Meyer, Parsons, & Dunne, 1990). It is further identified into two contrasting approaches i.e. deep approach and surface approach. A deep approach entails learners’ intrinsic motivation and interest to attempt to understand the meaning of the learning material and relate parts to each other, new ideas to previous knowledge or to personal meaningful context; whereas a surface approach is characterised by extrinsic motivation to acquire only sufficient knowledge to complete the task or pass the subject and thus, learners tend to memorise separate facts and/or view a particular task in isolation from other tasks and real life as a whole (Marton & Säljö, 1976; Biggs, 1985; 1987a; Biggs and Moore, 1993; Wong, Lin, & Watkins, 1996; Biggs, Kember, & Leung, 2001; Kember, Biggs, & Leung, 2004).
2.2 The Influence of ICT-Supported Learning Environment Perceptions on Approaches to Learning

There have been many studies reporting that significant relationships exist between students’ perceptions of learning context and approaches to learning. It was found that approaches to learning are influenced by different perceptions of students studying different subject areas (Ramsden, 1979). In this vein, accounting students were especially influenced by their learning context which perceived as being tensed up with pressure and demands from the professional accounting bodies and there was evidence that most of them adopt the surface approaches to learning (Eley, 1992; Gow, Kember, & Cooper, 1994; Sharma, 1997; Booth et al., 1999; Jackling, 2005; Lord & Robertson, 2006; Abraham 2006). On the other hand, deep approach to learning was found to be associated with perceived quality teaching support (Eley, 1992; Chan & Watkins, 1994; Lizzio, Wilson, & Simons, 2002), appropriate pedagogy which encourages independence, interaction, and inquiry (Eley, 1992; Abraham, 2006), and appropriate assessment (Abraham, 2006; Watty, Jackson, & Yu, 2010).

Furthermore, the ICT-supported learning environment is especially contributing to students’ perceptions of interactivity and involvement (Maor, 2000; Law, Lee, & Chow, 2002; de Lange, Suwardy, & Mavondo, 2003; Jebeile & Abeysekera, 2010; Premuroso, Tong, & Beed, 2011), inquiry and investigation (Basu & Cohen, 1994; Siragusa, 2002; Jones, Scanlon, Gaved, Blake, Collins, Clough et al., 2013), authenticity of learning (Basu & Cohen, 1994; Green, Reinstein, & Mc Williams, 2000; Murphy & Hoepnner, 2002; Marriott, 2004; Stanley & Edwards, 2005; Neal 2005), cooperation (Rumpagaporn, 2007), differentiation (Jebeile & Abeysekera; 2010) as well as the perceptions of teacher support (Rumpagaporn, 2007; Lillie & Wygal, 2011). These perceptions were found contributing to students’ deep approaches to learning. Meanwhile, the ICT-supported learning environment perceived by students to have replaced them by producing the accounting reports automatically (Green et al., 2000) and being a safety net for absence (Wells, de Lange, & Fieger, 2008) were associated to surface approaches to learning.

3. Methodology

This study was conducted through adopting the quantitative descriptive research design to explore students’ perceptions of ICT-supported learning environment and approaches to learning as well as their relationship. It was conducted through a survey where questionnaires were developed and distributed to the targeted respondents to collect the related information.

3.1 Subject

The subjects were 33 Form 4 (or Grade 10) students who are studying Principles of Accounting under an ICT-supported learning environment which is defined by Aldridge, Dorman, and Fraser (2004) as an environment where students could utilise various ICT tools to assist their learning. For example, using ICT to complete and submit assignments, search information, obtain notes, and conduct on-line discussion and communication with teacher and peers.

3.2 Measures

Students were asked to complete a questionnaire assessing their approaches to learning and perceptions of ICT-supported learning which was adapted from the Revised Two-Factor version of the Learning Process Questionnaire (R-LPQ-2F) (Kember, Biggs, & Leung, 2004), Technology-Rich Outcomes-Focused Learning Environment Inventory (TROFLEI) (Aldridge et al., 2004), and Course Experience Questionnaire (CEQ) (Wilson, Lizzio, & Ramsden, 1997). They were requested to respond to a five-point Likert scale for all the items in the questionnaire from “strongly disagree” to “strongly agree”. The questionnaire was back-to-back translated into the Malay language as most of the students in public secondary schools in Malaysia are proficient in this language.

The R-LPQ-2F was adapted to determine students’ approaches to learning. It consists of 22 items where both Deep and Surface Approach are measured by 11 items respectively. This instrument is chosen as it is specially designed to measure the approaches to learning of secondary
school students and it is a shorter version that is suitable for use as a classroom evaluation tool which was developed from the original version of Learning Process Questionnaire (LPQ) designed by Biggs (1987b). Moreover, this simple two-factor instrument was verified by the authors as valid and reliable with good psychometric properties. It contains scales that scored good Cronbach alpha values (Deep Approach = .82 and Surface Approach = .71) and exhibited good construct validity (Comparative Fit Index, or CFI = .967, Standardised Root Mean Squared Residual, or SRMR = 0.036).

On the other hand, students’ perceptions of ICT-supported learning environment were assessed through the instrument adapted from the Technology-Rich Outcomes-Focused Learning Environment Inventory (TROFLEI) (Aldridge et al., 2004) and Course Experience Questionnaire (CEQ) (Wilson et al., 1997). The TROFLEI provides the scales of Student Cohesiveness, Teacher Support, Cooperation, Involvement, Investigation, Task Orientation, and Computer Usage where each subscale was measured by seven to nine items; while the scale of Appropriate Assessment was developed from CEQ which is measured by five items. TROFLEI is renowned in terms of its reliability and validity for being extensively cross-validated in various countries such as Taiwan and Australia (Aldridge, Fraser, & Huang, 1999); and the United Kingdom, Canada, and Australia (Dorman, 2003). It is also advocated by the authors that the instrument showed good reliability where its scales yielded high Cronbach’s alpha values (ranging from .77 to .95) with sound construct validity (CFI=0.98). In addition, the Appropriate Assessment scale from CEQ was validated by the authors as it attained consistently good Cronbach’s alpha values ranging from .69 to .75. for three consecutive years of study. It also yielded high loadings for each item ranging from .56 to .85 which reflected good construct validity.

3.3 Procedures
The questionnaire was directly administered by the main author of the present study to all the subjects in order to ensure the genuineness of the data collection process and to obtain higher response rate. In addition, respondents were given prompt assistance should they encounter any difficulties in answering the questionnaire. Permission was first acquired from the school principal followed by discussion with the Principles of Accounting teachers on the suitability of date, time, and place for conducting the survey. The survey process started with a 10-minute briefing on the purpose of the research and instructions on how to answer the questionnaire.

4. Results
The first stage of this study employed the reliability test to assess the internal consistency of each subscale, followed by descriptive statistics in order to understand the overall status of respondents’ perceptions. Pearson correlation was used for analysis of relationships among the subscales in the second stage by reporting the inter-correlational matrix to provide an overview of the univariate relationships.

4.1 Reliability and Descriptive Statistics
Each subscale of the instrument achieved a high level of reliability, except for Surface Approach. Thus, a few items were dropped from this subscale in order to secure a satisfactory level of reliability. The reliability values i.e. Cronbach’s alpha values (α) together with descriptive statistics for each subscale are shown in Table 1.

All means are greater than 3.5, except the Surface Approach and Appropriate Assessment. It indicates that students generally have positive perceptions towards their learning environment and have low tendency to adopt surface approaches to learning but are more inclined towards using deep approach. However, the rather low mean value in Appropriate Assessment may reflect a more fact- and memory-oriented assessment for the subject of Principle of Accounting. In addition, the standard deviations for all subscales were less than 1 and this suggests that the scores tend to be very close to the mean scores.
Table 1: Reliability values and descriptive statistics of the subscales.

<table>
<thead>
<tr>
<th>Subscale</th>
<th>Mean</th>
<th>Standard Deviation</th>
<th>Items</th>
<th>Cronbach Alpha (α)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Approaches to Learning</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Deep Approach (DA)</td>
<td>3.83</td>
<td>.69</td>
<td>11</td>
<td>.91</td>
</tr>
<tr>
<td>Surface Approach (SA)</td>
<td>3.01</td>
<td>.66</td>
<td>6</td>
<td>.71</td>
</tr>
<tr>
<td><strong>Perceptions of ICT-Supported Learning Environment</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Student Cohesiveness (SC)</td>
<td>3.83</td>
<td>.73</td>
<td>8</td>
<td>.91</td>
</tr>
<tr>
<td>Teacher Support (TS)</td>
<td>3.79</td>
<td>.84</td>
<td>8</td>
<td>.92</td>
</tr>
<tr>
<td>Cooperation (CO)</td>
<td>4.06</td>
<td>.71</td>
<td>8</td>
<td>.93</td>
</tr>
<tr>
<td>Involvement (IV)</td>
<td>3.72</td>
<td>.64</td>
<td>8</td>
<td>.85</td>
</tr>
<tr>
<td>Investigation (IVT)</td>
<td>3.62</td>
<td>.76</td>
<td>7</td>
<td>.88</td>
</tr>
<tr>
<td>Task Orientation (TO)</td>
<td>3.88</td>
<td>.67</td>
<td>8</td>
<td>.89</td>
</tr>
<tr>
<td>Computer Usage (CU)</td>
<td>3.72</td>
<td>.94</td>
<td>9</td>
<td>.93</td>
</tr>
<tr>
<td>Appropriate Assessment (AA)</td>
<td>2.68</td>
<td>.89</td>
<td>5</td>
<td>.80</td>
</tr>
</tbody>
</table>

4.2 *Inter-correlation Matrix*

Table 2 shows the inter-correlation matrix for scores on the 8 subscales of Perceptions of ICT-Supported Learning Environment, Deep Approach, and Surface Approach.

Table 2: Inter-correlation matrix between Deep Approach, Surface Approach, and subscales of Perceptions of ICT-supported Learning Environment (N=33).

<table>
<thead>
<tr>
<th></th>
<th>DA</th>
<th>SA</th>
<th>SC</th>
<th>TS</th>
<th>CO</th>
<th>IV</th>
<th>IVT</th>
<th>TO</th>
<th>CU</th>
</tr>
</thead>
<tbody>
<tr>
<td>SA</td>
<td></td>
<td>.43*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SC</td>
<td>.43*</td>
<td></td>
<td>.03</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TS</td>
<td>.73**</td>
<td>.25</td>
<td></td>
<td>.57**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CO</td>
<td>.35*</td>
<td>.07</td>
<td>.66**</td>
<td>.50**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IV</td>
<td>.42*</td>
<td>-.05</td>
<td>.83**</td>
<td>.65**</td>
<td>.76*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IVT</td>
<td>.44*</td>
<td>.16</td>
<td>.43*</td>
<td>.47**</td>
<td>.72*</td>
<td>.64**</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TO</td>
<td>.56**</td>
<td>.23</td>
<td>.54**</td>
<td>.58**</td>
<td>.50**</td>
<td>.65**</td>
<td>.68**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CU</td>
<td>.51**</td>
<td>.03</td>
<td>.67**</td>
<td>.34</td>
<td>.34</td>
<td>.48**</td>
<td>.36*</td>
<td>.53**</td>
<td></td>
</tr>
<tr>
<td>AA</td>
<td>.02</td>
<td>-.10</td>
<td>-.29</td>
<td>-.04</td>
<td>-.42*</td>
<td>-.19</td>
<td>-.29</td>
<td>-.25</td>
<td>-.16</td>
</tr>
</tbody>
</table>

Notes: * = p < .05; **= p < .01

Deep Approach has significant positive relationships with all the subscales of perceptions, except Appropriate Assessment (r=.02). Meanwhile, Appropriate Assessment does not have significant relationships with most of the other subscales of perceptions. Furthermore, Teacher Support has the strongest positive relationship (r=.73) with Deep Approach which reflects the strong link between teacher’s role and students’ adoption of the deep approach. In addition, Teacher Support is significantly related to other subscales of perceptions such as Cooperation (r=.50), Involvement (r=.65), Investigation (r=.47), and Task Orientation (r=.58). It further suggests the importance of teacher’s role in the learning environment.

On the other hand, none of the subscales is significantly related to Surface Approach. This could suggest that the ICT-supported learning environment which forms students’ perceptions towards Student Cohesiveness, Teacher Support, Cooperation, Involvement, Investigation, Task Orientation, Computer Usage, and Appropriate Assessment is not associated to Surface Approach but more towards Deep Approach.

Furthermore, it is found that Computer Usage has significant relationships with many other subscales such as Student Cohesiveness (r=.67), Involvement (r=.48), Investigation (r=.36), and Task Orientation (r=.53). This suggests that ICT is related to many dimensions of the learning environment.
Lastly, it was found that many of the subscales of perceptions are strongly related, e.g., Student Cohesiveness and Involvement ($r=.83$), Involvement and Cooperation ($r=.76$), and Cooperation and Investigation ($r=.72$). It could suggest a phenomenon of multicollinearity which one should pay attention when conducting multiple regression analysis. Multicollinearity is an undesired statistical phenomenon in which two or more independent variables in a multiple regression model are highly correlated and will affect the accuracy of the model (Hair, Anderson, Tatham, & Black, 1995).

5. Discussion

From the findings, it reveals that the instrument measures for Surface Approach needed to be further refined as a few items had to be dropped in order to obtain a satisfactory but relatively lower reliability value compared to other subscales.

On the other hand, though there are many studies claiming that students adopt the surface approach for accounting by rote learning a set of rules (Eley, 1992; Beattie et al., 1997; Booth et al., 1999; Lucas, 2001), the present study preliminarily found that students demonstrated deep approach for learning Principles of Accounting. This phenomenon was found strongly associated to the perceptions of teacher support and it is consistent with many studies that found the significant relationship between teacher support and deep approach (Eley, 1992; Chan & Watkins, 1994; Lizzie et al., 2002), as well as teacher support and deep approach under the learning environment that supported by ICT (Rumpagaporn, 2007; Lillie & Wygal, 2011). Furthermore, the significant positive relationships between perceptions of teacher support and cooperation, involvement, investigation, or task orientation could suggest that the right pedagogy employed by a teacher which encourages independence, interaction, and inquiry is able to foster deep approach to learning which was proposed by Eley (1992) and Abraham (2006).

Moreover, this study also shows the significant positive relationships between perceptions of computer usage and many other perceptions that are simultaneously related to deep approach significantly such as student cohesiveness, involvement, investigation, and task orientation. These results are coherent with many studies which advocate the impact of these perceptions on deep approach to learning in the ICT-supported learning environment (Basu & Cohen, 1994; Maor, 2000; Law et al., 2002; Siragusa, 2002; de Lange et al., 2003; Jebeile & Abeysekera, 2010; Premuroso et al., 2011; Jones et al., 2013). Thus, the findings could suggest that ICT plays an important role in fostering deep approach to learning.

Finally, the perceptions of assessment which were found to be unrelated to deep approach may reflect that the perceived fact- and memory-oriented assessment could hardly be related to deep learning. The result is in contrast with the studies conducted by Abraham (2006) and Watty et al. (2010) which advocated that deep approach to learning is influenced by the perceptions of assessment.

6. Conclusion and Future Research

This preliminary study found that students demonstrated deep approaches to learning Principles of Accounting in an ICT-supported learning environment. In this vein, deep approach to learning was found to be positively related to various perceptions of ICT-supported learning environment, particularly the perceptions of teacher support and computer usage. It may reflect that a good teaching environment that employs appropriate pedagogy and technology could encourage students to adopt deep approaches to learning. This supports the concept of Technological Pedagogical Content Knowledge (TPCK) (Mishra and Koehler, 2006) which advocates that good teaching require an understanding of how technology constructively relates to pedagogy and content. Thus, technology is not just an “add on” in the teaching and learning process; rather, the skill and art of using technology and the context of its use are the key determinants for stimulating deep approach to learning.

On the other hand, more effort has to be put in to improve the assessment of Principles of Accounting as the current findings show a more fact- and memory-oriented assessment as perceived by students. As advocated by researchers, deep approach to learning is especially
important in the learning of accounting because most of the concepts must be mastered through understanding and not memorising (Sukumaran, 1991; Borthick & Clark, 1986). In this vein, students must see the general principles which organize all procedures of accounting into a coherent whole rather than rote learning the procedural rules. Therefore, appropriate assessment which evaluates students’ level of meaningful understanding is important for fostering deep approach to learning.

Further research has to be conducted to consolidate the research instrument, especially the items measuring Surface Approach, in order to obtain a more reliable result. On one hand, more samples have to be included for future studies to ensure the representativeness and generalisability of the sample. In addition, more variables have to be involved in predicting approaches to learning as suggested by Biggs (1985) because both the students’ personal and situational factors could influence learning approaches. Since the perceptions of ICT-supported learning environment which were studied in the present research are the situational factors, more personal factors have to be examined in future such as students’ academic abilities, prior educational experience, and personalities. Finally, by involving more samples and variables in this study, a more sophisticated analysis method such as Multiple Regression Analysis (MRS) or Structural Equation Modelling (SEM) (Bentler, 1980; 1983; Bollen, 1989) can be employed. The MRS is able to identify the best independent variable which can contribute to the deep or surface approach to learning, while SEM has the capability in estimating and testing hypothesised interrelationships among the observed and latent variables.

References


Technology-enhanced Chemistry Learning and Students' Perceptions: A Comparison of Microcomputer-based Laboratory and Web-based Inquiry Science Environment

Kulthida KAMTOOMa, Niwat SRISAWASDIb*

aScience and Technology Education Program, Faculty of Education, Khon Kaen University, Khon Kaen, Thailand
bDivision of Science, Mathematics, and Technology Education, Faculty of Education, Khon Kaen University, Khon Kaen, Thailand

*niwsri@kku.ac.th

Abstract: With current advancement of information and communication technology, researchers have indicated the necessity and challenges of developing effective instructional strategies by the use of technological tools to promote chemistry learning in school science. Consequently, the researchers have developed two kinds of technology-enhanced chemistry learning environment regarding the rate of reaction concepts: MBL-based guided-inquiry learning, as face-to-face instruction; web-based inquiry science environment (WISE), as online instruction. The purpose of this study was to investigate effects of both learning environment on students' perceptions. The participants were tenth graders; they are assigned randomly into two groups for interacting with different learning environment, MBL (N=37) and WISE (N=34). To assess students’ perceptions in the different groups, pre- and post-questionnaires were used. The results of the study (N=71) show that there was statistically significant difference between pre- and post-questionnaire for both groups. No significant difference was found between the MBL and WISE groups in terms of students’ perceptions. The result suggested that both technology-enhanced learning environment progressed students’ perceptions of learning and the progression was not indifferent. This finding highlight the benefits of face-to-face MBL and WISE to promote students’ perceptions of learning.

Keywords: MBL, WISE, chemistry learning, perception, rate of reaction

1. Introduction

To advance the practice of science education, technologies have become commonplace because of their potential of bringing about change in ways of teaching and learning (Srisawasdi, 2012). For this reason, in research and development of the student learning process in science, the effective use of technology in the classroom teaching process has become a common place in science-based education. In the community of science educator and educational researcher, it has been widely established that contemporary science learning environments should foster inquiry-based experiences and investigations. These inquiry-based scientific practices should take place in the laboratory in order to interact directly with naturally occurring phenomena (Srisawasdi, 2014; Pyatt & Sims, 2011). In recent years, many researchers have applied microcomputer-based laboratory into inquiry-based science education for teaching and learning (Friedler et al., 1990; Gunhaart & Srisawasdi, 2012; Liao & She, 2009; She & Liao, 2009) that can promote authenticity and application of scientific process and practice and facilitate concrete-to-abstract conceptualization of real-world phenomena (Pyatt & Sims, 2011). Likewise, Web-based Inquiry Science Environment (WISE) is an open-source digital learning platform that supports student inquiry in classrooms for support students visualization and understand in sub-microscopic levels (Chiu, Jennifer & Linn, 2011). Researchers reported that students could be provided effective learning experience of science and student can link the idea of three levels of chemistry
representation (i.e. macroscopic, sub-microscopic, and symbolic) (Chang & Linn, 2013). As abovementioned, both science learning environments were recognized as effective technological tool for inquiry-based learning process, and there is a plenty of challenge to think about how to use them in science teaching and learning pedagogically.

Accordingly, the aim of this study was to investigate students’ perceptions delivered in face-to-face MBL and online-mediated WISE. Specifically, the following questions were answered:

1. Do the students engaged in face-to-face MBL and online-mediated WISE learning environment perform significantly better by students’ perceptions?
2. Is there a difference between students’ perceptions between face-to-face MBL and online-mediated WISE learning environment?

Following the questions, it is hypothesized that both face-to-face MBL and online-mediated WISE will lead greater students’ perception after their participations, and there are no different on students' perceptions on experience with both learning environments.

2. Literature review

2.1 Microcomputer-based Laboratory (MBL)

Recently, In teaching and learning about of science-based education, computer technology has been widely used for modeling scientific knowledge structures or learning patterns, developing more in-depth and integrated understanding of concepts and process, enhancing the development of scientific skills, visualizing complex and dynamic scientific phenomena, promoting collaborative network in the community of learning for the construction of knowledge, and sharing of data, support access to a variety of information, support collecting various types of scientific data, test underlying theories through diagnostic or tutorial strategies, and enhance characteristics of inquiry as the way scientists work (Srisawasdi, 2014). In recent years, the reviewing of empirical evidence it is clear that microcomputer can improve the learning process of students in science education. Moreover, many researchers have applied microcomputer-based learning to inquiry-based science education for teaching and learning (Friedler et al., 1990; Gunhaart & Srisawasdi, 2012; Liao & She, 2009; She & Liao, 2009). The result of researchers shown that microcomputer-based inquiry learning can support student to learn science efficiency and make student can observe phenomena and understand in macroscopic level of representation.

2.2 Web-based Inquiry Science Environment (WISE)

WISE is the multimedia learning, in which the teachers can insert the lessons into web-based learning environment in order to aid and follow the learning results from students. WISE has four steps following Knowledge Integration (KI) framework (Chiu, Jennifer, & Linn, 2011): (i) eliciting ideas is to emphasize the native knowledge of the students; (ii) adding ideas is to add the new knowledge to the students; (iii) distinguishing ideas is to promote critical thinking of the students; and (iv) sorting out ideas is to rearrange the ideas of the students. In an addition, researcher reported impact of the WISE that students can link the idea of three levels of chemistry knowledge (i.e. sub-microscropic, macroscropic, and symbolic (Chang & Linn, 2013).

3. Method

3.1 Participants

The participants in this study were tenth-grade students (16-17 years old) in the Northeastern region of Thailand. The selection of samples performed randomly. They were divided into two groups which received different learning environment: WISE (N=34) and MBL (N=37) groups. Both groups were assigned to learn a chemistry lesson on rate of reaction. For online-mediated WISE class, the students were enabled to learning by self-regulated learning process with the WISE, which was designed the rate of reaction lesson followed KI framework. In contrast, the students in face-to-face MBL class were engaged into guided-inquiry learning process with the
support of MBL tools. However, two technology-based learning environments were equivalent in term of chemistry concepts on rate of reaction. The students had not had any formal education on rate of reaction before the study took place. Figure 1 illustrated information about participants and learning environment.

![Diagram of participants and learning environment](image)

**Figure 1.** Diagram of participants and learning environment

### 3.2 Learning Environments

#### 3.2.1 Face-to-face MBL

To engage student into face-to-face MBL, this study employed microcomputer-based laboratory (MBL) by Vernier and software technology. Vernier mini-LabQuest is a digital interface used to collect sensor data and then transfer data into Logger Pro software building graph and analysis application. For this study, students were provided opportunity to conduct an investigation of effects of reaction rate using carbon dioxide gas sensor. Figure 2 displays an example of hands-on MBL for guided-inquiry learning process in chemistry class.

![Example of hands-on guided-inquiry MBL on the effects of rate of reaction](image)

**Figure 2.** Example of hands-on guided-inquiry MBL on the effects of rate of reaction

As mentioned, guided-inquiry learning process was used to promote student's chemistry learning of rate of reaction. The simulation-based open inquiry with dual-situated learning model (Srisawasdi and Kroothkeaw, 2014; Srisawasdi and Sornkhatha, 2014) was applied for the face-to-
face MBL learning process in a manner of guided inquiry. Table 1 displayed an example of the learning process on rate of reaction concept.

Table 1: The six-stage guided-inquiry learning process for face-to-face MBL environment

<table>
<thead>
<tr>
<th>Stage</th>
<th>Learning process</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Pre-lab</strong></td>
<td></td>
</tr>
<tr>
<td>Open-ended inquiry question</td>
<td>Teacher provides an open-ended inquiry question: “How about rate of reaction if difference surface areas of sculpture react with acid rain is concentrated static?”</td>
</tr>
<tr>
<td>Scientific background/ information</td>
<td>Teacher induces collaborative discussion toward the definitions and pictorial diagram of acid rain reaction ; surface area ; CO$_2$ gas</td>
</tr>
<tr>
<td><strong>Lab practice</strong></td>
<td></td>
</tr>
<tr>
<td>Procedure/design</td>
<td>Teacher introduces the overall procedure of the laboratory, explains how to perform a predetermined experiment step-by-step, and guides how to record necessary data. Then, student conducted the experiment about acid rain (H$_2$SO$_4$) with sculpture (CaCO$_3$) interaction for collecting the experimental data from carbon dioxide gas when they change surface area of CaCO$_3$.</td>
</tr>
<tr>
<td>Data and result analysis</td>
<td>After the interacting with laboratory, students make a decision to analyze obtained experimental data from their own design and interpret it into results.</td>
</tr>
<tr>
<td><strong>Post-Lab</strong></td>
<td></td>
</tr>
<tr>
<td>Result communication</td>
<td>Students have to select the way to present, communicate, and discuss the meaning of data and experimental results to others, for an example, writing experimental question, experimental design, results, and discussion on a newsprint paper and then present to the class.</td>
</tr>
<tr>
<td>Conclusion</td>
<td>Students have to collaboratively make a relationship between each group results and then draw it into a conclusion as the best answer to the provided inquiry question. For an example, teacher induces each group make a conclusive answer by using an integration of their own and other results.</td>
</tr>
</tbody>
</table>

In this learning process, student will collaborative work together in small groups of three to five members. This pedagogy begins with an open-ended driving question targeted to alternative conceptions about effects of rate of reaction commonly found in students. To assist the process of hypothesis generation addressed the question, essential scientific backgrounds are provided to students. Then, students are required to perform generating testable hypotheses. Information about scientific process and procedures were given to guide them performing experimentation with MBL tools. During experimenting with the MBL, instructor came to probe them with a series of formative assessment question. When they finished the experiment, all group has to communicate findings among groups by creating a chart of results and explaining findings. Finally, instructor induces students into a forum for drawing a conclusion based on evidence and collaborative explaining the result of hypotheses testing. Figure 3 displays an example of students’ interaction in the context of hands-on guided-inquiry MBL.
3.2.2 Online-mediated WISE

WISE curriculum materials is an online-mediated environment which features powerful, interactive scientific visualizations to illustrate unobservable phenomena such as chemical reactions at the atomic level, large-scale phenomena such as climate change, and phenomena that happen quickly such as collisions. The online environment logs student interactions and responses to embedded assessments. WISE offers teachers grading and classroom management tools to monitor student progress, provide feedback to students, and personalize instruction. In this study, the researchers created a chemistry learning unit on rate of reaction on WISE system by emphasizing the interplay between macro-, sub-micro-, and symbolic representations of chemistry. The learning unit was constructed following Knowledge Integration (KI) framework. Figure 4 displays screen interface of WISE system for the chemistry learning unit of rate of reaction.

(A) Elicit Ideas: asking students what happen in this picture? for eliciting their existing knowledge of the students

(B) Add idea: giving student necessary information for adding new related knowledge to existing ideas
(C) Distinguishing idea: using of simulation and animation to promote students' scientific thinking and reasoning  
(D) Sorting out idea: Asking students a question, “What factors affect the weathering of architecture from acid rain?” to rearrange conceptual ideas

Figure 4. An illustrative examples of screen interface of the chemistry learning unit of rate of reaction in online-mediated WISE following KI framework, from (A) to (D)

In this learning process, student will collaborative work together in small groups of two or three members. The teacher gave project ID to students in order to access the WISE and then facilitate students' group learning in the class and monitor progression of learning process during their participations. Finally, instructor induces students into a forum for drawing a conclusion based on evidence and collaborative explaining the rate of reaction phenomena. Figure 5 displays an example of students' interaction in the context of online-mediated WISE.

Figure 5. An illustrative example of classroom learning activity through online-mediated WISE.

3.3 Instruments

A 16-item Likert-scale questionnaire was developed to use in this study for examining students’ perceived ease of use (3 items), perceived usefulness (3 items), and perceived satisfaction (3 items), obtained from Barzilai & Blau (2014), while an instrument by Cheng (2014) was used to measure perceived learning (4 items), and enjoyment (3 items). To develop a Thai version of the questionnaire, the original English version was translated identically in Thai language. Two experts were recruited to identify communication validity of the items. On each item, respondents were assigned to rate how much the respondent agree with into five scale, from 1-strongly disagree to 5-strongly agree. The instrument had been established validity and reliability. Table 2 displayed sample items of the questionnaire used in this study and its reliability.
Table 2: Sample items of the Likert-scale questionnaire measured students' perceptions

<table>
<thead>
<tr>
<th>Dimension</th>
<th>Sample items</th>
<th>α</th>
</tr>
</thead>
<tbody>
<tr>
<td>WISE</td>
<td>• It is easy for me to learn how to use WISE.</td>
<td>0.737</td>
</tr>
<tr>
<td></td>
<td>• I can easily accomplish what I need to do in WISE.</td>
<td></td>
</tr>
<tr>
<td>MBL</td>
<td>• It is easy for me to learn how to use MBL.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• I can easily accomplish what I need to do in MBL.</td>
<td></td>
</tr>
<tr>
<td>Perceived ease of use</td>
<td>• WISE can help me learn more effectively.</td>
<td>0.842</td>
</tr>
<tr>
<td></td>
<td>• WISE can improve my course performance.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• MBL can help me learn more effectively.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• MBL can improve my course performance.</td>
<td></td>
</tr>
<tr>
<td>Perceived usefulness</td>
<td>• I feel comfortable to use WISE.</td>
<td>0.774</td>
</tr>
<tr>
<td></td>
<td>• I enjoy the experience of using WISE.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• I feel comfortable to use MBL.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• I enjoy the experience of using MBL.</td>
<td></td>
</tr>
<tr>
<td>Perceived satisfaction</td>
<td>• I learned a lot from the WISE.</td>
<td>0.803</td>
</tr>
<tr>
<td></td>
<td>• The WISE added to my knowledge.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• I learned a lot from the MBL.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• The MBL added to my knowledge.</td>
<td></td>
</tr>
<tr>
<td>Perceived learning</td>
<td>• I enjoyed the WISE.</td>
<td>0.754</td>
</tr>
<tr>
<td></td>
<td>• Interacting the MBL was pleasant.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• I enjoyed the MBL.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Interacting the MBL was pleasant.</td>
<td></td>
</tr>
</tbody>
</table>

3.4 Data Collection and Analysis

Students were investigated perceptions by the 5-point Likert-scale questionnaire before giving face-to-face MBL and online-mediated WISE intervention for 10 minutes. Both learning environments, students participated to interact with them for 60 minutes. After the instruction, students were administered by the same questionnaire again as post-test. The statistical data techniques selected for analyzing students' perceptions was repeated-measures MANOVA in SPSS 21.0.

4. Results and Discussion

The results for the repeated-measures MANOVA indicated no significant main effect for group (Wilks' lambda=0.878, F(5, 65) = 1.798, p=0.126, partial η² = 0.122). There was no significant difference on students' perceptions between face-to-face MBL and online-mediated WISE groups. The univariate results on two group revealed none of the five subscales on perception reached a statistical significance between both groups. That is, both face-to-face MBL and online-mediated WISE group performed indifferently with regard to perceived ease of use, perceived usefulness, perceived learning, perceived satisfaction, and enjoyment. Also, there was no significant interaction effect between group and time (Wilks' lambda=0.934, F(5, 65)=0.923, p=0.472, partial η² = 0.066). This means that both face-to-face MBL and online-mediated WISE has similar effects on students' perceptions. However, there was a significant main effect for time (Wilks' lambda=0.567, F(5, 65)=9.940, p<0.001, partial η² = 0.433). This suggests that, on average, students' perceptions have changed over the learning environments. Univariate analyses of variances (ANOVA) on each subscale were conducted as follow-up tests to the one-way MANOVA. The results of the univariate test for face-to-face MBL and online-mediated WISE groups are summarized in Table 3.

Table 3: The students' subscale means of perceptions by group and time and univariate MANOVA by time

<table>
<thead>
<tr>
<th>Subscale</th>
<th>MBL</th>
<th>WISE</th>
<th>F</th>
<th>Sig.</th>
<th>η²</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Pre-test</td>
<td>Post-test</td>
<td>Pre-test</td>
<td>Post-test</td>
<td></td>
</tr>
</tbody>
</table>

111
As can be seen in Table 3, The univariate MANOVA on the five subscale scores of perceptions were significant differences across time, from pre-test to post-test. The univariate results revealed a significant effect on perceived ease of use \( (F_{1,69} = 43.741, p < 0.01, \eta^2 = 0.388) \), perceived usefulness \( (F_{1,69} = 34.515, p < 0.01, \eta^2 < 0.333) \), perceived learning \( (F_{1,69} = 31.124, p < 0.01, \eta^2 = 0.311) \), perceived satisfaction \( (F_{1,69} = 42.716, p < 0.01, \eta^2 = 0.382) \), and enjoyment \( (F_{1,69} = 25.807, p < 0.01, \eta^2 = 0.272) \). According to aforementioned results, the overall result suggested that the increase of students’ perceptions regarding perceived ease of use, perceived usefulness, perceived learning, perceived satisfaction, and enjoyment from the pre-test to post-test was homogeneous both face-to-face MBL and online-mediated WISE group. This result consistent with findings that students perceived online learning environment easy to use, to learn, and to accomplish what need to be done (Barzilai & Blau, 2014). In an addition, it confirm with previous studies (Voogt, Tilya and van den Akker, 2009) that students were satisfied with hands-on MBL experience for getting more investigative and open-ended learning, and user friendly.

5. Conclusion

The results revealed that there is no interaction effect on group (face-to-face MBL and online-mediated WISE) and time (pre- and post-test) regard perceived ease of use, perceived usefulness, perceived learning, perceived satisfaction, and enjoyment. This means that the increasing of students' perceptions did not depend on different technology-enhanced learning environments e.g. face-to-face MBL and online-mediated WISE. This finding highlights the importance of the need for considering effective instructional design and environment to enhancing chemistry teaching and learning. For the chemistry class, both face-to-face MBL and online-mediated WISE group could be used to facilitate teaching and learning of chemical concepts and support emotional practice in chemistry education. For the next study, the authors have a plan to implement both learning environments for investigating whether if it would be more beneficial to combine WISE and MBL activities than to use them separately in teaching rate of reaction.

Acknowledgements

This work was financially supported by the Institute for the Promotion of Teaching Science and Technology (IPST) and Graduate School, Khon Kaen University. The author would like to express gratefully acknowledge to Faculty of Education, Khon Kean University, for supporting this contribution.

References


The Design of Instructional Scaffolds to Facilitate Online Project-Based Learning

Chun-Ping WUA*, Ching-Chiu YEHB, Shu-Ling WC & Hao Jie YONGC

ADepartment of Educational Technology, TamKang University, Taiwan
BDepartment of Educational Technology, TamKang University, Taiwan
CGraduate Institute of Human resource and eLearning Technology, National Hsinchu University of Education, Taiwan
*cpwu303@gmail.com

Abstract: The study designed scaffolds to facilitate online project-based learning. The question-generation, worked examples and peer-assessment were adopted and transformed into the scaffolds. The integration of the scaffolds into the project process was exemplified in a distance course, entitled as instructional design. Thirty college students participated in the study for 18 weeks. The Moodle as well as the Adobe-connect learning system were adopted to engage learners with the contents, scaffolds and team collaboration. Data analysis found that the scaffolds facilitated participants’ engagement with the learning contents and team interaction. Empirical significance of the study as well as suggestions for instructional implementation and future study are provided.

Keywords: Online project-based learning, Worked examples, Scaffolds, Peer-assessment

1. Introduction

Creating instructional products with virtual teams to effectively engage learners in active knowledge creation by maximizing the merits of advanced interactive learning technology is very complex (Perez & Emery, 1995). Instructional designers’ cognitive abilities to integrate multiple domains of knowledge and collaboration skills to bring together the diverse expertise of each team member at the distant are highly demanded. The online project-based learning strategy is frequently adopted to engage student designers in deliberately practicing the above-mentioned process (Tynjala, 1998).

However, the inherent openness of project-based learning environment might introduce even more complex cognitive tasks which might increase student designers’ cognitive loads. Without systematic guidance to facilitate learners in constructing essential knowledge bases, student designers often fail to execute relevant domain knowledge in analyzing the project scenarios. Moreover, they are often not able to efficiently manage project progress and exchange ideas among virtual team members, which might further result in the incompleteness of project tasks. Therefore, design appropriate scaffolds to provide support on a conceptual level and virtual team interaction are essential.

This study aims to design the scaffolds for implementing project-based learning in the distance course of instructional design. More specifically, the design of the worked example, question generation strategy and peer-assessment as the scaffolds is proposed. Students’ engagement in the learning process and interaction among peers are investigated.

2. Literature Review

Project-based learning, which is rooted in the idea of learning by doing, enables student designers to practice solving diverse design problems, synthesize theories and skills, and further construct their knowledgebase (Tynjala, 1998). Its effect is influenced by three design elements: (1) The relevancy of the project scenario with the real-world design problems: The more the project
simulates the real-world problem, the more complex of the project scenario will be. The complexity of the project brings student designers more challenge; meanwhile, it might also enhance their curiosity and intrinsic goal orientation (Bednar, Cunningham, Duffy, & Perry, 1992; Cobb, 1994; Karagiorgi, & Symeou, 2005). (2) Systematic guidance: The project usually presents learners with ill-structured design problems, requiring multiple thinking dimensions, which in turns result in student learners heavy cognitive loads. Without adequate and timely guidance, learners might easily give up their projects especially while working with peers at the distant on the virtual project. Moreover, student designers, lacking of the experience in working on ill-structure problems, tend to think of the design problem from single perspective and over-emphasize on the project product instead of the project process. (3) Team facilitation: Student designers’ abilities to construct their cognitive structures are enhanced by social interaction with peers, which includes asking questions, explaining their reasoning for the design ideas, listening to peers’ ideas and reconciling differences between their ideas and the ideas of others (O’Donnell & King, 1999). However, collaboration levels in the virtual projects are difficult to control, which leads to the lack of clarity of individual responsibilities and leadership and low intensity of interaction among the virtual group members. Therefore, a sound design of scaffolds is essential to engage student designers in reasoning through the problem from multiple perspectives, provide challenge and encourage more creative design ideas (Tynjala, 1998) and facilitate them in working on virtual projects.

Two essential types of scaffolds to facilitate online project-based learning are conceptual scaffolds and interaction scaffolds. The support of the conceptual scaffold using the worked example could effectively reduce a novice designer’s cognitive load in analyzing and solving a new ill-structured problem (Sweller, 1998; Sweller, 2005; Turner, Meyer, Cox, Logan, DiCintio, & Thomas, 1998; van Merriënboer, Kirschner, & Kester, 2003). However, the reduction of the extraneous cognitive loads does not necessarily initiate their devoting more working memory to exploring the complex structures represented in the worked examples and the underlying principles and knowledge associated with the given problem. In other words, an effective worked example should engage student designers in executing working memory to simulate expert thinking process and reason through the solution path. Moreover, the cognitive load brought by the project tasks will change according to a designer’s expertise. A designer, who is able to transfer the learned knowledge and skills into schema, could efficiently focus on the design/solution path and underlying knowledge structures presented in the worked example. Too much guidance might bring them extra cognitive load (Kalyuga, Ayres, Chandler, & Sweller, 2003; Kalyuga & Sweller, 2004), Therefore, the conceptual support should gradually fade out with student designers’ development of the expertise.

Scaffolding effects of question-generation and peer-assessment strategies can be reasoned based on social learning and constructivism (Yu, & Liu, 2005; van Gennip, Segers & Tillema, 2009). On one hand, the effects of these two strategies in enhancing learners’ comprehension of the contents are evidenced in many empirical studies. On the other hand, the high cognitive abilities demanded by the collaborative question-generation task might contribute to high density of the interaction within virtual teams and encourage them to share and construct knowledge. Moreover, the peer-assessment process enables student designers to observe peers’ products, compare their own design solutions with those designed by others, which might lead to modification of their products and their own cognitive structures.

Several research gaps after in-depth literature review were identified: First, the linkage between the worked examples and novice instructional designers’ engagement in the activity of constructing their knowledge base in the distant learning environment remains less known. Second, the scaffolding effects of question-generation and peer-assessment strategies on cognitive abilities can be reasoned based on social learning and constructivism; however, their effects on the team interaction require more empirical evidence. Last, but the most importantly, reflection is importance in developing instructional designers’ competency. Majority of the research emphasizes on the role of “self” in the reflection process (Bilinski, 2002; Johns, 1995; Schmidt, 2004). However, argued from the collective intelligence perspective, that peer assessment and dialogue could enhance more deliberate and collaborative reflection is reasonably anticipated (Sluijsmans, Moerkerke, van Merriënboer, & Dochy, 2001; Sluijsmans, Brand-Gruwel, & van
Merriënboer, 2002). Therefore sound design of these scaffolds emphasizing on peer collaboration in the distant course is worthy of more efforts.

3. Research Method

3.1 Research Design

Thirty college students, who registered in the distant course, entitle as instructional design, participated in the study for 18 consecutive weeks. The asynchronous features of the Moodle learning system and the synchronous features of the Adobe-Connect were adopted in this study to engage learners with the learning activities, contents, scaffolds etc. To ensure that participants possessed the fundamental skills of using the communication technologies embedded in the learning systems, a training session on integrating the technologies for project communication and the follow-up question-generation and peer assessment with hands-on activity was arranged.

For the duration of the study, students attended face-to-face sessions at the first two weeks to get familiar with the learning path of the overall course, their peers and the learning technologies. The learners were randomly assigned to six teams. In the following 16 weeks, the class moved on to an asynchronous format. During the weekly online learning activity, students watched the video explaining the learning contents individually, followed by a team mission to be completed. Each member within the team took turns on leading the team to accomplish the team mission. In addition to the weekly team mission, each team was required to an instructional product as their course project. At the last class, they moved on to individually assess the instructional product developed by other teams.

The design blueprint of the scaffolds was developed and implemented. Two data sources are collected to explore the impact of the scaffolds. First, the frequency of participants’ accessing the systems was used to evidence levels of their engagement with the course content and the scaffolds. Second, the number and quality of the posting within each team were presented and analyzed weekly.

4. Results and Conclusions

4.1 Design of the Scaffolds

Two types of scaffolds to facilitate online project-based learning were designed and exemplified in Table 1. The purposes and implementation of the scaffolds associated with the four stages of online project-based learning were described.

<table>
<thead>
<tr>
<th>Learning stage</th>
<th>Aim</th>
<th>Scaffold</th>
<th>Levels</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.Project initiation</td>
<td>Associate the course progress with the project process</td>
<td>Learning Map</td>
<td>Individual</td>
</tr>
<tr>
<td></td>
<td>Introduce the project</td>
<td>Project Descriptor</td>
<td>Group</td>
</tr>
<tr>
<td>2-1.Knowledge building</td>
<td>Ensure the knowledge base for the project</td>
<td>Question-Generation</td>
<td>Individual</td>
</tr>
<tr>
<td>2-2.Team Building</td>
<td>Enhance team interaction</td>
<td>Question-Assessing</td>
<td>Group</td>
</tr>
<tr>
<td>3-1.Problem Analysis</td>
<td>Exemplify the analysis process</td>
<td>Worked Examples</td>
<td>Group</td>
</tr>
<tr>
<td>3-2.Research /Development</td>
<td>Exemplify the research/development process</td>
<td>Worked Examples</td>
<td>Group</td>
</tr>
<tr>
<td>4.Solution Evaluation</td>
<td>Seek feedback</td>
<td>In-field testing</td>
<td>Group</td>
</tr>
</tbody>
</table>
1. Project initiation stage: Learning maps were designed to match the project tasks and requirement with the weekly learning contents. This scaffold does not only enable individual learner to link learning contents with the learning outcomes but also scaffold learners to plan their learning path. The second scaffold introduced is the project descriptor. This study created an animation to present different project scenarios and the e-tutor within the animation guided the learners to explore the information inherent in the scenario. Teams were asked to reach consensus in selecting one scenario as their term project based on their exploration of the scenarios.

2. Knowledge and team building stage: The scaffold of question-generation was implemented to ensure individual’s comprehension of the course contents by asking each learner to generate one multiple-choice question. The weekly leaders of the teams have to facilitate their team discussion to modify the generated questions.

3. Problem analysis, research and development stage: The scaffold of worked example with e-tutor exemplified experts’ reasoning process and guided learners to think of multiple perspectives and synthesize the required domain knowledge.

4. Solution evaluation stage: Each team conducted a field-testing of their design products and revised before in-class peer review. During the peer review process, each individual evaluated the assigned product according to the given criteria and provided written recommendation. After the peer-assessment, the team members were required to read and interpret the comments and write a reflection on product improvement.

Learners’ engagement with the learning content is evidenced in 798 times of visiting the content weekly. More specifically, the frequencies of each learner interacting with the contents during the period of introducing the scaffolds of learning maps, project descriptors, question-generation and worked-examples are 26.6, 10.4, 13.2, and 10.53, respectively.

<table>
<thead>
<tr>
<th>Learning Maps</th>
<th>Project Descriptor</th>
<th>Question-Generation</th>
<th>Worked-Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Weekly hits</td>
<td>798</td>
<td>312</td>
<td>396</td>
</tr>
<tr>
<td>Weekly hits per learner</td>
<td>26.6</td>
<td>10.4</td>
<td>13.2</td>
</tr>
</tbody>
</table>

Team interaction is revealed in 743 postings supplied by the learners and the 959 times of visiting the posting during 18 weeks. More specifically, the weekly conversations within teams occurring during the period of introducing question-generation scaffolds range from 8 to 39 iterations while the team conversations occurring during the period of introducing the worked example scaffold range from 1 to 25 iterations (Please see Table 3).

<table>
<thead>
<tr>
<th>No. of subjects</th>
<th>Question-Generation</th>
<th>Worked-Example</th>
<th>Peer-Assessment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Team1</td>
<td>6</td>
<td>20</td>
<td>7</td>
</tr>
<tr>
<td>Team2</td>
<td>4</td>
<td>12</td>
<td>2</td>
</tr>
<tr>
<td>Team3</td>
<td>4</td>
<td>28</td>
<td>8</td>
</tr>
<tr>
<td>Team4</td>
<td>5</td>
<td>8</td>
<td>1</td>
</tr>
<tr>
<td>Team5</td>
<td>5</td>
<td>12</td>
<td>2</td>
</tr>
<tr>
<td>Team6</td>
<td>6</td>
<td>39</td>
<td>25</td>
</tr>
<tr>
<td>Total</td>
<td>30</td>
<td>119</td>
<td>45</td>
</tr>
</tbody>
</table>

In addition to the number of postings, the team interaction should also account for the frequency of reading the postings, which can represent inherent cognitive interaction within the
team. Table 4 presents the weekly frequency of reading the postings within teams during the period of introducing question-generation scaffolds, ranging from 128 to 536 times. More interestingly, the weekly frequency of reading the comments provided by their peers during the peer-assessment activity ranges from 92 to 217.

Table 4. The weekly frequency of reading postings and replied postings within the team

<table>
<thead>
<tr>
<th>No. of subjects</th>
<th>Question-Generation</th>
<th>Worked-Example</th>
<th>Peer-Assessment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Team 1</td>
<td>6</td>
<td>300</td>
<td>86</td>
</tr>
<tr>
<td>Team 2</td>
<td>4</td>
<td>233</td>
<td>243</td>
</tr>
<tr>
<td>Team 3</td>
<td>4</td>
<td>298</td>
<td>82</td>
</tr>
<tr>
<td>Team 4</td>
<td>5</td>
<td>128</td>
<td>21</td>
</tr>
<tr>
<td>Team 5</td>
<td>5</td>
<td>150</td>
<td>33</td>
</tr>
<tr>
<td>Team 6</td>
<td>6</td>
<td>536</td>
<td>348</td>
</tr>
<tr>
<td>Total</td>
<td>30</td>
<td>1644</td>
<td>613</td>
</tr>
</tbody>
</table>

4.2 Conclusion

The study designed scaffolds to engage student designers in conducting instructional projects with virtual teams. First, the results imply that the designed scaffolds could facilitate student designers’ engagement with constructing their knowledge base in the distant learning environment. Second, the scaffolding effects of question-generation and peer-assessment strategies on the team interaction are also evidenced.

The findings have important empirical significance as well as implications for research on online project-based learning. First, the study substantiated the effect of question-generation on learners’ mastery of the content Second, the worked-example scaffold which simulates experts’ reasoning process, facilitates students in synthesizing knowledge associated with solving the given problem, which is the key factor to the success of project-based learning. Finally, the team interaction was enhanced by the scaffolds. In-depth and systematic analysis of the conversations between interacting parties and learners perceptions toward group dynamic via interaction process analysis is recommended for future study.

References


GeMA-ICT Learning Effectiveness in Improving Student Mathematical Ability

Sigid Edy PURWANTO*, Wahidinb & Aidlyah Novian NISYAHc

*Mathematics Education, UHAMKA University, Indonesia
bMathematics Education, UHAMKA University, Indonesia
cMathematics Education, UHAMKA University, Indonesia

sigidmath@yahoo.co.id

Abstract: This study examines and measures the effectiveness of using GeMA-ICT learning methods (Games, Manipulatives, and Activities with Information and Communication Technologies) on the students’ mathematical abilities. GeMA-ICT is a new method that resulted from the theoretical studies of the Mathematics Education Study Program UHAMKA research team. Mathematics has abstract objects such as abstract facts, abstract concepts, abstract operations, and abstract principles. The abstract objects in mathematics education sought to be easily understood by students, by presenting them through concrete objects, props math, math games and math activities supported by the use of ICTs. Props, games, and activities developed with the help of ICTs are a blend of media and learning methods that can visualize math concepts. A set of concrete objects are intentionally drafted, designed, manufactured, assembled, and used for instilling or developing concepts or principles in mathematics. With GeMA-ICT methods, things that can be presented in the form of abstract models such as concrete objects can be viewed, held and manipulated so that they can be easily understood. In addition to presenting math games, learning math is made fun for students and learning activities are turned into a process of investigation and exploration of further mathematical concepts.

Keywords: GeMA-ICT, props math, math games, math activities

1. Introduction

Extant research has shown that effective learning seldom happens in a scenario where the teacher dominates the process and reduces the students to passive participants. Whereas learning ought to be made easy for students, lots of teachers are still making it difficult for them because they do not reinvent their classroom procedures according to the patterns of modern teaching approaches. Classroom activities such as, “one-way lectures” and “take-home-assignments” only teach students to perform certain symbolic procedures as well as to work but not to think. This is what characterizes most mathematics classrooms – students are often confined to a scenario where they simply watch the teacher solve mathematical problems on board and at the end of the class they would be given assignments to take home and solve them in worksheets according to the pattern that the teacher used, (Turmudi, 2008).

Another characteristic of the traditional teaching approach is that teachers emphasize on curriculum attainment targets rather than targeting at mastery of the material. Many math teachers teach by simply following routines that are critically ineffective. This eventually wears them out because the process is not only tedious to the teacher but also damaging to students’ interest (Sobel and Maelisky, 2004). Furthermore, in the traditional classroom setting, teachers tend to use chalk and talk method often more than not. This was caused by several possibilities: 1) schools already have props but are not using them optimally; 2) school do not have props; 3) school have had adequate props but are not good enough (Asyhadi, 2005).

To reinvent the current classroom scenario that still retains the traditional characteristics, and enhance the quality of national education; research must focus on teachers and educational facilities and infrastructure. Learning in the traditional classroom setting does not give room for maximal innovation and creativity and it hinders the opportunity of using a wide variety of
methods and media. As a result, learners gain minimal knowledge little understanding of the content of the lessons. Hence, the learning process becomes unattractive and weakening to the national education standards. Adding up all these together, the standard of the competence of graduates becomes significantly undermined. Hence, there is need to test pilot the usage of contextual and humanistic learning approaches in today’s mathematics classrooms. Accordingly, the attention of governments and experts of mathematics education in many countries is quite crucial at this point in time, so as to improve the mathematical ability of students and equally champion gigantic efforts for overcoming students’ passiveness in mathematics classrooms.

Results in NAEP study show that students are still having difficulties when faced with problems that require reasoning and problem-solving abilities. Mathematics learning of the present day has not been able to develop students' mathematical problem solving abilities optimally. This causes the learning conditions of students to be only able to resolve problems only in accordance with the examples given by the teacher. But when students are given another problem that is similar but not the same as the examples once given, they find it difficult or even impossible to handle. It was observed in NAEP study that the success rate of students in solving problems dropped dramatically when the context is replaced with things that are not known to the students (Suherman et al, 2003).

However, one of the new offers that can activate students' learning in today’s Mathematics classrooms is learning with GeMA-ICT (Games, Manipulatives, and Activities with Information and Communications Technologies). This lesson allows students to be actively involved through hands-on activity. GeMA-ICT is an emphasis on the learning activity of students in the manipulation of concrete objects by exploiting the use of math games, math props, and activities with the help of ICTs. GeMA-ICT encourages students to learn facts, skills, concepts and theories through manipulative activities with concrete objects, models, props, or math games. GeMA-ICT can increase the desire to learn; enhance learning by doing; and enhance the application of scientific problem solving through making analyses and evaluations. GeMA-ICT insists on ensuring that mathematical principles are found, generalized, and proven. This learning approach takes off the abstract nature of mathematics, making it interesting through the integration of games and a variety of other ICT activities.

In this approach, students’ thinking phases are made concrete and real, so as to allow for the execution of visual-kinesthetic activities a variety of ICT tools. On the whole, GeMA-ICT is a good option for overcoming the dull routine of learning in Maths classrooms. It is outrightly appropriate learning for enhancing students' mathematical abilities. Therefore, the purpose of this study is to determine the differences that are likely to be found in mathematics learning outcomes of students using GeMA-ICT method and students using expository method.

2. Literature

2.1 GeMA-ICT

GeMA-ICT is an abbreviation for Games, Manipulatives, and Activities with Information and Communications Technologies. Author deliberately adopted GeMA-ICT in order to obtain the equivalent word that can represent the learning activities of students who will be using the games, props, and math activities with ICTs.

When referring to the percentage of the amount that can be remembered, GeMA is a very important lesson. Johnson and Rising in Ruseffendi (2006) posited that, “learners can remember about a fifth of the heard, half of which is seen, and three-quarters of the done”. But in learning Mathematics, the manipulation of concrete objects is very important; yet in most cases the concrete objects used are ordinary local materials.

For learning to be more meaningful teachers should as much as possible avoid dominating the process. Literature has shown that in a teacher dominated classroom scenario, learning tend to be target-oriented-mastery in nature, and such learning has only proved successful in short term given competitions, while in the long run it doesn’t provide the child with the desired problem solving ability.
Abstract mathematical objects have facts, concepts, principles and principles of operation, which are as well abstract. It seems that students do easily come to grasp with lessons that are backed up with an orientation of concrete phenomena more than those that are not.

The works of Piaget, Bruner and Dienes have supported the claim that, manipulation of concrete objects is an important activity in mathematics learning. In GeMA, students solve problems, explore mathematical concepts, formulate and experiment with mathematical principles, and make mathematical discoveries through manipulation of concrete objects that represent abstract mathematical ideas.

GeMA learning follows the principle of learning by observing and doing; and it starts from the concrete to the abstract, as is the case with inductive method. Hence, students learn the objects and then generalize while ignoring the special nature of mathematical abstraction. It can attract learners to abstract mathematics. According to Ernest in Turmudi (2008) that learning mathematics is first and foremost is active, with students learning through games, activities, investigations, projects, discussion, exploration, and discovery.

Students can learn facts, skills, concepts, postulates, or theories through manipulating concrete objects, models, props, or math games. Hence, GeMA can increase students’ desire to learn, since it is based on the principle of learning by doing; and it also gives room for the students to appreciate and apply the scientific method of problem solving. However, application of learning aids such as laboratory equipment and other media becomes necessary for the implementation of learning with GeMA. But while the design of learning outside the classroom is directed to how students can play while learning, GeMA can be operationalized through the implementation of games, props, and math activities.

2.2 Math Games

Basically the students would love the games and puzzles, because play is indeed a world of children (Turmudi, 2008). Imam Al-Ghazali said, "Playing around for a child is something that is very important but banning him from playing around will turn him off and disturb his sense of belonging, intelligence and general rhythm of life". Children will find it easier to learn arithmetic by means of handing out apples to them than by abstract examples. Congruently, play is seen as a natural activity in helping the child to gain experiences, develop creativity and determination to succeed.

If a mathematical concept is presented through play, the understanding of the concept is expected to be steady, because learning in this way is in a natural pattern, which is in accordance with the child's instincts. Hence, the learning process is a psychological process, not a logical process. Therefore, the patterns should not be mere series of knowledge that have been previously defined in form of a mechanical process; but rather, through play, the students construct their mathematical patterns (Hudoyo, 1985).

Math game is a fun activity that can support the achievement of learning goals in mathematics cognitive, affective, and psychomotor domains. Math games help students to memorize basic facts, find the arithmetic operations and improve numeracy skills, as well as gain more understanding and problem solving ability (Ruseffendi, 2006).

Learning mathematics by games and puzzles was also emphasized by Turmudi (2008), as an approach for motivating and giving fun to both students and teachers alike in the learning process. This is important because games and puzzles have been widely recognized as a way of inspiring students to mathematical literacy. Ernest in Turmudi (2008) claims that games teach math effectively due to: 1) Provision of reinforcement and skills practice, 2) Provision of motivation, 3) Assistance, acquisition and development of mathematical concepts, and 4) Development of problem-solving strategies. Posamentier and Stepelman in Turmudi (2008) presented an analogy between game strategy problems solving strategy in the following table:
Table 1: The Comparison of Games Strategy and Problem Solving

<table>
<thead>
<tr>
<th>Games Strategy</th>
<th>No</th>
<th>Problem Solving Strategy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Read the rules</td>
<td>1</td>
<td>Read the rules</td>
</tr>
<tr>
<td>Understand the rules</td>
<td>2</td>
<td>What is given and what to look for?</td>
</tr>
<tr>
<td>Develop a plan</td>
<td>3</td>
<td>Write the equation</td>
</tr>
<tr>
<td>Work the plan</td>
<td>4</td>
<td>Solve the equation</td>
</tr>
<tr>
<td>If you win, smile; if not, think about why it lost</td>
<td>5</td>
<td>Check the answer</td>
</tr>
</tbody>
</table>

Moerlands and Makkink (2003), reported that the play activity could help resolve the problem of the "unknown" number. Their study revealed improvement in child’s learning outcome by 40%, from a mean baseline (2.6) to 3.7. Congruently, results in this study indicate that the learning using games makes abstract problems seem ordinary (Armanto, 2003). Meanwhile Benko & Maher (2006) have reported significant improvements in students’ oral, written ability and physical representation ability to level 7 after learning through games that use dices.

2.3 Props Mathematics

Props are teaching media that contains or carries the characteristics of the concept being studied. A set of concrete objects are designed, manufactured, assembled, or prepared deliberately to help embed or develop concepts or principles in mathematics. With props, things that can be presented in the form of abstract models such as concrete objects that can be seen, held and distorted are used for teaching, so that lessons taught can be easily understood.

Since the 1950s until the 1970s, research into the use of props in teaching mathematics has been going on, and not less than 20 summaries of such researches have been recorded. Among these is the popular summary of Higgins and Suydam in 1976 (Lithanta, 2009), which among other things concluded as follows: 1) In general, the use of visual aids in the teaching of mathematics was successful or effective in promoting student achievement. 2) Approximately 60% vs 10% of the students sampled showed a convincing success of learning with the props than without. 90% of students that learnt with props recorded yet the same magnitude of 90% in their learning outcomes above those students that did not use props. 3) Manipulation of the visual aids is important for elementary students at all levels. 4) Only a little evidence was found showing that manipulation of props is only manageable at lower learning levels.

Slamet, Soenarto, and Wahidin (2008), reported that the ability to compute and factorize quadratic equations and increase students’ learning outcomes becomes easy when lessons are administered with props, AEM (Block al-Khawarizmi). Congruently, studies have shown that learning with games could serve the needs of students at all levels; and even weak students could easily manipulate concrete rectangular objects using their prior knowledge of the broad concept of the rectangle, (Dienes AEM). Hence, a problem is often best solved (understood) by using sketches, folded pieces of papers, pieces of strings, or other simple props available. Invariably, strategic use of props can make the situation real to the students so as to motivate them to learn faster and better. Therefore, manipulation of geometry models can be a way of help the problem solving process as well as an activity for innovation, (Sobel & Maletsky, 2004).

Mathematics learning activities wherever possible involve all the senses of the students, especially hearing, seeing, and touching. In this case props bridge gap between the abstraction process and apprehension. In addition, by using the props, the child can be helped to find a strategy to solve problems. This is done by allowing the child to describe the problem in a simple concrete pattern; construct his or her own knowledge and understanding of the issues for the purpose of develop problem-solving strategies (Triyana, 2004).

2.4 Activities

In learning with games, activities are equated with experiments in a way that present lessons for students to conduct experiments and prove their own experiences and the things they learnt. In this
case, students are given the opportunity to experience for themselves, learn on their own, following a process that allows them to observe objects, analyze issues and draw conclusions, (Djamarah and Zain, 2006). Through this practicum process, the students are able to discover facts and truths in the form of conjectures and theorems by themselves.

Learning mathematics through practical applications or hands-on experience is an activity within the framework of the invention and principles of mathematical concepts. The process is helpful in improving students’ ability to explore, investigate, and the draw conclusions through physical activity, as well as through mental and emotional engagements, (Krismanto, 2003). For the full geometry of the material abstraction, hands-on mathematics is still a necessary experience for improving students’ learning outcomes.

With the mock objects (models) or concrete objects that are deliberately prepared to further stimulate the minds of students in constructing their own understanding, there are more elements of practical work on learning experiences for students to use the knowledge they gained (according constructivism) and not to solely depend on how their teachers teach math. This is an advocacy for a paradigm shift in teaching mathematics.

From the analysis of data from the 1996 NAEP test, two samples of countries involving 15,000 students mentioned that the rate of 8 students whose teachers actively taught through the process of learning activities generated employments in mathematics achievement levels of more than 70% (Crawford, 2001).

Vui (2006-2007) reported that the goal of good practice in teaching mathematics is to help students make meaning of the contents and skills taught in the lessons (what is known) and the process involveds (what is done). Good Practicum must balance between the content and the process of learning problem solving skills, because the two are entirely different aspects of the knowledges which the students must be equipped with, all their lives. When teachers use manipulative materials in teaching mathematics, they discover that their students are more active in learning. Students enjoy learning mathematics with dynamic models or motions. Teachers must learn how to create new mathematical models of problematic situations and prepare good manipulative materials for students to deal with. Students may also be inspired to build their own questions and activities.

2.5 Using ICTs

Using ICTs in this study refer to the use of Microsoft power point and GeoGebra Software. The use of Microsoft power point should be simplified in a way that will make it a preferred attraction to many. Currently, Microsoft PowerPoint is widely used by teachers in delivering course materials. Most teachers consider the use of Microsoft power point as an effective and efficient tool for practical learning.

In the use of GeoGebra many interesting things can be encountered and difficult mathematical problems such as in calculus lessons are easily resolved with the GeoGebra media software. Hence, apart from calculus problems, other mathematical problems, such as solving line equations, vectors, angles, algebras, geometry and many others can be easily resolved by the use of the GeoGebra media software. In addition, GeoGebra is designed to facilitate its use in an interactive way.

3. Research Method

3.1 Research Instruments
This research was conducted among the 150 Junior High Schools at Kramat Jati, East Jakarta on January 20 until February 4, 2014. This study uses a quasi-experimental design. In a quasi-experimental, the subjects are not taken at random but researchers do accept existing subjects. The use of quasi-experimental research is based on the consideration that there are classes that were not considered previously formed at random grouping of individuals who would disturb the Teaching and Learning Activities at schools.

The population in this study were students of 150 Junior High Schools, that comprised up of 8 eighth grade classes. This research used two classes as a sample, each of which has the same
characteristics. The first class was used as an experiment class, and it was taught by using the GeMA-ICT. The other class was used as a control class, and it was taught without using the GeMA-ICT.

Data was drawn from the results of learning mathematics scores among 71 students sampled. The research instrument used to measure students' mathematics learning outcomes was designed with multiple choice questions having four (4) alternative answers. The validity of the instrument was measured using the point biserial correlation formula, by which the instrument earned as much as 20 valid questions. Reliability of the instrument was measured by using the Kuder - Richardson (KR-20) formula, by which the instrument a very high reliability degree of $R_{11} = 0.836 > r_{table} = 0.329$.

3.2 Data Analysis

Students' mathematical ability tests were analyzed by using the inferential normality test, homogeneity test and t-test, to see the effectiveness of a given learning GeMA-ICT method.

4. Result

4.1 Student Mathematical Ability

The descriptive statistics research data obtained from the study of mathematics students in the experimental class (i.e. the class taught with GeMA-ICT method) is presented in Table 2 below:

<table>
<thead>
<tr>
<th>Data</th>
<th>Maximum Score</th>
<th>Student Learning Outcomes Math Scores</th>
</tr>
</thead>
<tbody>
<tr>
<td>35</td>
<td>20</td>
<td>$Y_{\min}$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>9</td>
</tr>
</tbody>
</table>

The descriptive statistics research data obtained from the study of mathematics students in the control class (the class that was taught without using GeMA-ICT method) is presented in Table 3 below:

<table>
<thead>
<tr>
<th>Data</th>
<th>Maximum Score</th>
<th>Student Learning Outcomes Math Scores</th>
</tr>
</thead>
<tbody>
<tr>
<td>36</td>
<td>20</td>
<td>$Y_{\min}$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>7</td>
</tr>
</tbody>
</table>

4.2 Results of Testing Data Analysis Requirements

Using the Lilliefors test, it was concluded that the experimental and the control class data derived from the sample had a normal distribution.

Using the Fisher test, it was concluded that the samples of the two classes of experimental class taught using GeMA-ICT method and control class taught without using GeMA-ICT method have the same conditions or homogeneous variance.

4.3 Results of Hypothesis Testing

Through the calculation of the average grade, experimental and control classes obtained $t = 3.06$ and $t (0.95; (69)) = 1.67$ with a significance level $\alpha = 0.05$ and (df) = 69. Since $t = 3.06 > 1.67 = t (0.95; 69)$, this means the rejection of the research hypothesis $H_0$ which stated that there are significant differences in mathematics the learning outcomes of students taught with GeMA-ICT
method and those taught without using GeMA-ICT method in the 150 Jakarta Junior High Schools.

5. Conclusion

GeMA-ICT methods improve students’ mathematics learning outcomes owing to students’ usage of manipulative activities, props and games in this study. The process also allows students to discuss with each other in the group. On the whole, the GeMA-ICT method improves students’ learning outcomes better than other methods of teaching mathematics.

Acknowledgements
We would like to thank University of Muhammadiyah Prof. Dr. HAMKA of Jakarta for supporting this study.

References
THE EFFECTIVITY OF ASSOCIATION PICTURE MEDIA APPLICATION TOWARD THE KATAKANA LETTER READING COMPREHENSION OF GRADE TEN STUDENTS OF SMK MANAJEMEN (MANAGEMENT VOCATIONAL SCHOOL) JAKARTA

Restoe NINGROEM* & Endy Sjaiful ALIM**

*State University of Jakarta, Indonesia
**University of Muhammadiyah Prof. DR. HAMKA (UHAMKA), Jakarta, Indonesia
*restoe.ningroem@yahoo.com, **esaiib@yahoo.com

Abstract: The purpose of this research is to acknowledge the effectivity of association picture media application toward the students’ katakana letter reading comprehension. The hypothesis proposed that the comprehension of students’ katana letter reading, which are exposed to association picture in class is more effective compared to the students which aren’t exposed with such media. The research was conducted over grade ten students of SMK Manajemen Jakarta. The research used quasi experiment. The sample was taken from 50 students, which categorized 25 students as experimental class and 25 students as controlled class. The data collection technique used an instrument of katakana alphabet reading comprehension test. The requirement test of data analysis used in this research was Lilliefors Test for normality and Fisher-test for homogeneity, it was revealed that the class was normal and homogeneous. Hypothesis assessment used t-test which showed tcount = 7.67 ≥ 1.67 = ttable that Ho was rejected. It is revealed from the research that association picture media is effective to increase student’s comprehension in reading katakana letter.

Keywords: Learning Media, Association Picture, Reading Katakana.

1. INTRODUCTION

Indonesia has had bilateral cooperation with other countries for a long time. In engaging cooperation with other country, it is necessary to comprehend the language, because it is the sound symbol system that articulated (produced by speech device) which is used by community members to cooperate, interact, and self recognition (TPKB, 2005).

One of the important evidences of language in engaging cooperation with Japan has been shown by the presence of Japanese curriculum in middle schools in Indonesia. Acquiring skill in Japanese is a basic step and access for our people to study, get knowledge, understand their culture, and have employment in Japan, a modern and developed country. A country that has rich natural resources, various cultures, and positive life values that deserve to learn from.
In order to have skill in Japanese, it is necessary to read researches of Japanese literature first hand. One of the components of the Japanese writing system is Katakana letter.

In reality, students find it difficult to memorize Katakana compare to Hiragana, because Hiragana letter appears more in the text book. Students’ obstacle in learning Katakana is revealed to be psychologist one, such as: interest, attitude, confidence, and intelligence.

One of the efforts to overcome the issue is to use media in teaching Katakana letter. According to Hamalik, using media in the learning process can stimulate new will and interest, arouse motivation and eagerness in learning, and even affects psychologically to students. It can also ease students to comprehend, presents data interestingly and reliably, interprets data and information easily (Rohani, 1997).

One of teaching media that eases students in learning Katakana is association pictures, the benefit of such media can deliver messages, idea, etc. That involves little verbal languages but, leaves a deeper impression (Arsyad, 2003). With association picture media, learners will have long term memorization of katakana letters because association makes words easy to be stored in the memory and to be retrieved if needed since they already accustom to the words (Sameto, 2003).

2. LITERATURE REVIEW

2.1 Learning Media

Association of Education and Communication Technology (AECT) in USA identifies media as all forms and channels which are used to deliver messages or information (Arshad, 2005). Gerlach and Ely (1971) give the definition of media broadly and narrowly. Broadly, it means every individual, material, or event that gives the student opportunity to gain knowledge, skill, and manner. Therefore the media doesn’t simply mean things or objects, it can also mean an individual and a learning event. Teacher, text book, school environment can be the media. While narrowly, media is non-personal media that teacher use as people in charge of the learning process to achieve goals. That way media tend to be looked as graphical tools, photograph, or electronic tools to grasp, reshapes visual or verbal information (Roshidi, 2009).

According to Hamalik (1986) educational media is something that can be digested by the senses, shape and things which are visual and audio, used as a means of communication in the learning process, as an aid in learning process and related to teaching method. Picture is a crucial visual aid and available everywhere. Crucial because it can replace verbal words, concretes the abstract. Pictures enable people to grasp the idea or information contained clearly, vividly more than words can say (Munadi, 2007). As the Chinese saying, pictures speak more than a hundred words (Adiman, 2009).

2.2 Association

The association is relating one event to another event, between someone and other people, which is considered as a related series and interconnected to each other. In learning a foreign language, we must develop association as a mean that allows us to memorize words easily and brings full picture and any circumstances that embedded in that word (Sameto, 2003). The ability to associate in learning foreign language needs to be developed according to particular events.

Association is available to be developed for all life aspects and circumstance that related to our life (Sameto, 2003). Memorization in conventional ways is by applying left brain hemisphere by memorizing and repeating materials, while the association is done by applying right brain hemisphere to memorize symbols or pictures which is easier and quicker to absorb. From descriptions and benefits of association elaborated above. It can be concluded that the association picture media discussed in this research are pictures that serves as a noun or verb that resembles a Katakana letter as a series of meaning that interconnected to each other to smoothen long term memorization of Katakana letter shapes.
2.3 Reading Comprehension

Anderson (Syamri, 2011) defined reading as uttering written language symbols. While A.S.Broto and Syamri (2011) stated that reading is the uttering of sound symbols.

Aspects of reading:
1. Mechanical Skills, which is considered in the lower order. This aspect covers:
   a. Identifying letter form
   b. Identifying linguistic parts (phoneme, word, phrase, clause pattern, sentence, etc)
   c. Identifying relation or correspondence of spelling and sound pattern
   d. Low level of reading speed.
2. Comprehension skills, which is considered in the higher order. This aspect covers:
   a. Comprehend meaning definition.
   b. Comprehend meaning.
   c. Evaluation or assessment (content, form)
   d. Flexible reading speed, that adjustable to particular situation.

From both important reading aspects in reading, it can be inferred that reading letter is in the early introduction of mechanical skills, which takes place before one do comprehension skills.

2.4 Katakana Letters

Katakana letters are formed from straight and sketchy lines or chokusenteki. These straight and stroke lines that differ them from hiragana (Ang, 2005).

![Katakana Letter](image)

The word "kata" means partial, incomplete, or separated. The word "katakana" means "kana separation", as katakana script which originated from more complicated kenji component (Japan: an Illustrated Encyclopedia, 1993).

In koujien dictionary, the function of katakana (Izuru, 1998) is:

"Genzai de wa omo ni gairaigo ya giongo nado no hyouki ni mochiiru ", Which means in the contemporary time, generally katakana letter is used for foreign language writing and onomatope.
Although *katakana* and *hiragana* letters are in the group of *kana* letter, but the function is different. *Katakana* letter is used in the writing of Japanese words which comes from absorption of foreign languages, foreign countries, animals, foreigner’s name, plants, and foreign cities (Ang, 2005). *Katakana* also often (not always) used for transcription of Japanese company, e.g: Suzuki becomes スズキ, and Toyota becomes トヨタ. *Katakana* also serves as reinforcement, specially on signs, advertising and billboard. For example, seems common to see koko コ コ (here), gomi ゴミ (waste), or megane メガネ (glasses). Reinforcement using *katakana* in sentences is sometimes done by writers.

*Katakana* sometimes serves as *hiragana* replacer or as *furigana* to give utterance a word that is written in Latin letter, or for foreign word, which is written in *Kanji* for meaning, kanji but meant to be read as the origin form.

*Katakana* sometimes also used to indicate words that speech in foreign accent or unusual, in foreign character, robot, etc. For example, in *manga*, foreign character speech or robot can be represented by konnichiwa コンニチワ (hello) not *hiragana* which is more typical こんにちは.

---

**Figure 2.** Association Picture Media sample

---

**Figure 3.** Association Picture Media sample

---

2.5 Critical Framework

Since it is important to improve *katakana* reading comprehension of grade ten students of SMK, and the benefit of using association picture media is to ease students in memorizing the shape of *katakana*.
letter in a long term, then it is assumed that association picture media is suitable to be implemented in the learning process of *katakana* letter. So, it can be inferred that association picture media is effective to improve *katakana* letter reading comprehension.

3. **METHOD AND SAMPLING**

This research applied quantitative approach using Quasi Method. Sampling was treated by the technique of cluster sampling. Two classes of grade ten students of SMK Manajemen Jakarta are determined as experimental and controlled classes. Instruments used for the research consisted of a set of test and questionnaire as additional data. There were three kind of test forms; multiple choice, translation of katakana into latin, and vocabularies matching.

4. **FINDING AND DISCUSSION**

The research shows the end result of the comprehension of students in reading *kata*na letter as follow:

**Figure 4 : Experimental Class Score**

Figure 4 shows that the score of the comprehension of reading *kata*na letter of the Experimental Class, from 10 students the scores are in the span of 94-99 and the highest score is 100, and the average score is 90.5.

**Figure 5 : Controlled Class Score**

Figure 5 shows that the score of the comprehension of reading *kata*na letter of the Controlled Class, from 10 students the scores are in the span of 59-71 and the highest score is 97, and the average score is 61.2.

Looking at the test result the comprehension of reading *kata*na letter of experimental class and controlled class, it is revealed that the average score of experiment class is higher than the score of control class. Looking at the data analysis, it is noted that nul hypothesis (Ho) succesfully rejected and research hypothesis (HI) was successfully accepted. This finding was revealed by t-test formulation, tcount was found at 7.64 and t table which came from degree of freedom at 48 with significant level at 0.05 was 1.67. After research, t count was found to be at 7.638. The requirement of H0 rejection and HI acceptance was th> t. Since t count was higher than t table (7.638 > 1.67), so Ho is rejected and HI was accepted which means "the application of association picture media to increase
reading comprehension of grade ten students of SMK Jakarta Manajemen is more effective compared to the opposite approach”.

After a comparative test conducted using t-test, it showed that $t_{count} = 7.67 \geq 1.67 = t_{table}$ with the level of significant 0.05, it was assumed that the comprehension of reading *katakana* letter from students who were exposed with association picture media was better than those who weren't exposed by the media.

5. CONCLUSION

Based on the research of data analysis evaluation and all data that the writer received for this research, a conclusion reached, that applying educational media in learning process of Japanese will help to reach learning goals, since one of the benefits of using educational media is to make learning easy.

In this research association picture media is concluded to be effective to improve students’ comprehension of reading *katakana* letter. The special feature of association picture media that presents pictures that resembles and inter-relates *katakana* letters eased student to memorize *katakana* letter. Besides that, using association picture media, learning process becomes less boring since the students were exposed with attractive pictures and little verbal elaboration (in written or speech only).

References


133
http://Apository.upi.edu/operator/upload/s_prs_981119 chapter2.pdf. accessed on 23 April (2011)
An Online Survey: Studying the Antecedents of Technology Use through the UTAUT Model among Arts and Science Undergraduate Students

Priscilla MOSES*, Tiny Chiu Yuen TEY, Phaik Kin CHEAH, Timothy TEO and Su Luan WONG

Universiti Tunku Abdul Rahman, Malaysia
University of Macau, ChinaSAR
Universiti Putra Malaysia, Malaysia
*priscilla@utar.edu.my/prismo6@yahoo.com

Abstract: Students need to be well-equipped with the necessary information, understanding, capabilities, skills and awareness to learn a subject and simultaneously to optimize the use of technology. For that reason, this research studied the antecedents of students’ technology use through Unified Theory of Acceptance and Use of Technology (UTAUT) model. Besides that, this paper sought to explore whether there is a difference between Arts and Science undergraduate students in terms of technology use. There were 38 Arts and 30 Science undergraduate students who participated in this online survey. Based on the independent-samples t-test, there was no significant difference (t(66) = .558, p = .579) found in terms of technology use among the Arts (M = 5.772, SD = .653) and Science (M = 5.661, SD = .980) students. The magnitude of the differences obtained was very small. Therefore, the findings of the study suggest that both the Arts and Science students make use of the technology regardless of their major.

Keywords: Antecedents, technology use, UTAUT model, online survey

1. Introduction

One of the key players to successfully integrate technology into the education system is the student. They need to be well-equipped with the necessary information, understanding, capabilities, skills and awareness to learn a subject and simultaneously to optimize the use of technology. Technology is not only a medium to deliver or receive knowledge, but it also acts as a vehicle that helps students to travel along the pathway to prepare them for their future. According to Godin and Goette (2013), students who graduate these days regardless of their major, need to have the capabilities to work in a global marketplace and use whatever technology that is needed to work virtually.

There is also an increasing need for educators to incorporate technology in teaching and learning in universities in Malaysia. Identifying the differences would help the educators to address the challenges faced by Arts and Science students in the teaching pedagogy. A study conducted among undergraduates found that there was no significant difference in the overall scores between undergraduates from the Arts and Science disciplines in an ICT literacy course (Wong & Cheung, 2012). However, Liberal Arts and Business students were found to use less applications in their laptops compared to students who are in the Science disciplines (Percival & Percival, 2009). Despite the greater use of applications, another study reported that there was no significant difference in problem solving skills between Arts and Science students (Williamson, 2011).

Hence, this study sought to study the antecedents that influence the students’ technology use through the Unified Theory of Acceptance and Use of Technology (UTAUT) model as a research framework. An online survey was employed to measure six constructs: Performance Expectancy (PE), Effort Expectancy (EE), Social Influence (SI), Facilitating Condition (FC),
Behavioural Intention (BI), and Use Behaviour (UB). Additionally, this study tested whether field of study (Arts and Science) plays a role among undergraduates’ technology use.

2. Literature review

2.1 The UTAUT Model

Quite a number of theoretical models have been suggested to facilitate the understanding of factors impacting the user acceptance and usage behaviour of information technology. These models are universally used to predict and explain individuals’ behaviours towards technology acceptance (Dulle & Minishi-Majanja, 2011), such as Theory of Reasoned Action (TRA), Theory of Planned Behaviour (TPB), Technology Acceptance Model (TAM), Motivational Model (MM) and so forth were incorporate in the area of perceived ease of use as a determinant of acceptance (Liu & Kostiwa, 2007). Among all the models, Technology Acceptance Model (TAM) is one of the most widely applied and influential models in explaining information technology adoption behaviour (Venkatesh & Davis, 2000).

Venkatesh, Morris, Davis, and Davis (2003) formulated the more recent instrument, Unified Theory of Acceptance and Use of Technology (UTAUT) model in which they included the eight well-known models - Motivational Model (MM), Theory of Planned Behaviour (TPB), Technology Acceptance Model (TAM), Theory of Reasoned Action (TRA), Model of PC Utilization, Innovation Diffusion Theory, Combined TAM-TPB, and Social Cognitive Theory. The UTAUT model does not only describes the main individual-level factors that influence technology acceptance, but the possibilities that would limit and amplify the influence of these factors (Venkatesh & Zhang, 2010). The credibility of the UTAUT model is established in explaining a large portion of variance in the user behaviour intention towards the use of technology (Venkatesh & Zhang, 2010) and it has been validated outside the origin where it was first proposed (Teo & Noyes, 2012). Besides, there are four constructs in UTAUT model which play key roles as direct determinants of user acceptance and usage behaviour - performance expectancy, effort expectancy, social influence and facilitating conditions (Venkatesh et al., 2003).

The UTAUT model presents three direct determinants (see Figure 1) to assess behaviour intention towards the use of technology (performance expectancy, effort expectancy, social influence), two direct determinants of technology use (behaviour intention and facilitating conditions), and four contingencies (age, gender, experience and voluntariness) affecting behaviour and/or intention towards the use of technology (Venkatesh & Zhang, 2010). However, the four contingencies in the UTAUT model were excluded in this research because they are moderating variables which affect the relationship between the determinants and technology use behaviour; while the focus in this research is to examine the direct factors that affect the undergraduates’ technology use behaviour (Baron & Kenny, 1986; Brown, Dennis, & Venkatesh, 2010). In the present study, the researchers also sought to explore if field of study plays a role in technology use.

Over the last decade, the UTAUT model has been widely used to examine technology use in educational context, especially in e-learning and mobile learning (Cruz, Boughzala, & Assar, 2014; Lin, Lu, & Liu, 2013; Thomas, Singh, & Gaffar, 2013). According to Cassidy et al. (2014), technology evolution has impacted education as students’ exposure to technology has increased dramatically including computer, mobile software, electronic gadgets and social networks. As Cassidy and her colleagues reported, students’ technology use for academic purpose, such as the use of e-reader, has doubled in four years. Hence, technology evolution has also contributed to ubiquitous use and access in education. As said by Godin and Goette (2013), future studies should be conducted to examine the virtual learning and technology acceptance with the intention to comprehend better on how to prepare the students to collaborate virtually in a global environment by incorporating these determinants.
3. Purpose of the study

This study aims to achieve the following objectives: (a) to explore the antecedents that explain the students’ technology use through the UTAUT Model, and (b) to test whether there is a significant difference in technology use between Arts and Science undergraduate students.

4. Methodology

4.1 Participants

The sample consists of 68 Arts and Science undergraduates from a private university located in peninsula Malaysia. There were 18 males (26.5%) and 50 females (73.5%) who participated in the online survey. Table 1 provides a summary of the undergraduates’ ages. The undergraduates’ age ranges from 20 to 27. The mean age of the participants is 22.26 with standard deviation of 1.39.

<table>
<thead>
<tr>
<th>Age</th>
<th>Frequency</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>8</td>
<td>11.8</td>
</tr>
<tr>
<td>21</td>
<td>9</td>
<td>13.2</td>
</tr>
<tr>
<td>22</td>
<td>21</td>
<td>30.9</td>
</tr>
<tr>
<td>23</td>
<td>23</td>
<td>33.8</td>
</tr>
<tr>
<td>24</td>
<td>5</td>
<td>7.4</td>
</tr>
<tr>
<td>27</td>
<td>2</td>
<td>2.9</td>
</tr>
<tr>
<td>Total</td>
<td>68</td>
<td>100.0</td>
</tr>
</tbody>
</table>

In addition, Table 2 shows the undergraduates’ majors. There were 38 (55.9%) Arts undergraduates from Faculty of Arts and Social Science while the remaining 30 (44.1%) were Science undergraduates from the Faculty of Science.

<table>
<thead>
<tr>
<th>Major</th>
<th>Frequency (f)</th>
<th>Percentage (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Faculty of Arts and Social Science</td>
<td>38</td>
<td>55.9</td>
</tr>
<tr>
<td>Faculty of Science</td>
<td>30</td>
<td>44.1</td>
</tr>
<tr>
<td>Total</td>
<td>68</td>
<td>100.0</td>
</tr>
</tbody>
</table>
4.2 Research Instrument

The online survey was adapted from the UTAUT model instrument which was developed by Venkatesh, et al. (2013). In this research, the online survey was designed using Google Form. There were two sections in the online questionnaire with a total of 28 items. The respondents filled in their age and major in the first section and clicked on an appropriate option (7-point likert scale from “Strongly Disagree” to “Strongly Agree”) for the second section. Subsequently, their responses were recorded and submitted to a Web server, which was used to administrate the online survey.

Besides that, a reliability analysis was executed for the scales using Cronbach’s Alpha. As summarised in Table 3, all of the scales tested in the UTAUT constructs were reliable as each computed statistic showed a value above .70 ranging from .70 to .96. The Cronbach’s Alpha value of the questionnaire with 28 items was reported to be .95.

Table 3: Instrument Reliability

<table>
<thead>
<tr>
<th>Scales</th>
<th>Number of Items</th>
<th>Cronbach’s Alpha</th>
</tr>
</thead>
<tbody>
<tr>
<td>Performance Expectancy</td>
<td>4</td>
<td>.85</td>
</tr>
<tr>
<td>Effort Expectancy</td>
<td>4</td>
<td>.88</td>
</tr>
<tr>
<td>Social Influence</td>
<td>4</td>
<td>.79</td>
</tr>
<tr>
<td>Facilitating Conditions</td>
<td>5</td>
<td>.88</td>
</tr>
<tr>
<td>Behavioural Intention</td>
<td>5</td>
<td>.96</td>
</tr>
<tr>
<td>Use Behaviour</td>
<td>6</td>
<td>.70</td>
</tr>
</tbody>
</table>

5. Results and Discussion

A descriptive statistical analysis describing the antecedents of undergraduates’ technology use is presented in Tables 4, 5, 6, 7, 8 and 9. Table 4 is a summary of descriptive analysis for the undergraduates’ performance expectancy. As demonstrated in Table 4, the statistics suggest that the undergraduates perceive technology as an effective tool that enhances their studies and task accomplishment and productivity. Nevertheless, it appears that the undergraduates tend to be more neutral with respect to the perception that using technology will improve their academic grades.

Table 4: Descriptive Statistics for Performance Expectancy (PE) (n= 68)

<table>
<thead>
<tr>
<th>Questionnaire Item</th>
<th>1 Strongly Disagree</th>
<th>2 Disagree</th>
<th>3 Slightly Disagree</th>
<th>4 Neither Agree Or Disagree</th>
<th>5 Slightly Agree</th>
<th>6 Agree</th>
<th>7 Strongly Agree</th>
<th>Mean</th>
<th>Std. Dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>S11: People who influence my behaviour think that I should use technology.</td>
<td>0 (0%)</td>
<td>3 (4.4%)</td>
<td>14 (20.6%)</td>
<td>21 (30.9%)</td>
<td>18 (26.5%)</td>
<td>8 (11.8%)</td>
<td>4 (5.9%)</td>
<td>4.38</td>
<td>1.23</td>
</tr>
<tr>
<td>S12: People who are important to me think that I should use technology.</td>
<td>2 (2.9%)</td>
<td>5 (7.4%)</td>
<td>8 (11.8%)</td>
<td>21 (30.9%)</td>
<td>15 (22.1%)</td>
<td>13 (19.1%)</td>
<td>4 (5.9%)</td>
<td>4.43</td>
<td>1.43</td>
</tr>
</tbody>
</table>
SI3: The administration of this university has been helpful in the use of technology.

SI4: The university has supported the use of technology.

Table 5 provides the descriptive analysis for undergraduates’ effort expectancy. It shows that the undergraduates are confident in using technology as they believe that learning and operating technology is easy and understandable for them.

Table 5: Descriptive Statistics for Effort Expectancy (EE) (n= 68)

<table>
<thead>
<tr>
<th>Questionnaire Item</th>
<th>1 Strongly Disagree</th>
<th>2 Disagree</th>
<th>3 Slightly Disagree</th>
<th>4 Neither Agree Or Disagree</th>
<th>5 Slightly Agree</th>
<th>6 Agree</th>
<th>7 Strongly Agree</th>
<th>Mean</th>
<th>Std. Dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>EE1: My interaction with technology would be understandable.</td>
<td>1 (1.5%)</td>
<td>3 (4.4%)</td>
<td>2 (2.9%)</td>
<td>15 (22.1%)</td>
<td>21 (30.9%)</td>
<td>20 (29.4%)</td>
<td>6 (8.8%)</td>
<td>5.00</td>
<td>1.28</td>
</tr>
<tr>
<td>EE2: It would be easy for me to become skilful at using technology</td>
<td>0 (0%)</td>
<td>1 (1.5%)</td>
<td>4 (5.9%)</td>
<td>13 (19.1%)</td>
<td>19 (27.9%)</td>
<td>19 (27.9%)</td>
<td>12 (17.6%)</td>
<td>5.28</td>
<td>1.22</td>
</tr>
<tr>
<td>EE3: I would find technology easy to use.</td>
<td>0 (0%)</td>
<td>2 (2.9%)</td>
<td>4 (5.9%)</td>
<td>16 (23.5%)</td>
<td>18 (26.5%)</td>
<td>17 (25.0%)</td>
<td>11 (16.2%)</td>
<td>5.13</td>
<td>1.28</td>
</tr>
<tr>
<td>EE4: Learning to operate technology would be easy for me.</td>
<td>0 (0%)</td>
<td>3 (4.4%)</td>
<td>4 (5.9%)</td>
<td>13 (19.1%)</td>
<td>21 (30.9%)</td>
<td>17 (25.0%)</td>
<td>10 (14.7%)</td>
<td>5.10</td>
<td>1.30</td>
</tr>
</tbody>
</table>

Table 6 represents the undergraduates’ perceptions on social influence towards their technology use. The descriptive statistics suggest that the undergraduates are neutral in terms of their perception that important people around them and the university administration might affect their technology use.
Table 6: Descriptive Statistics for Social Influence (SI) (n= 68)

<table>
<thead>
<tr>
<th>Questionnaire Item</th>
<th>1 Strongly Disagree</th>
<th>2 Disagree</th>
<th>3 Slightly Disagree</th>
<th>4 Neither Agree Or Disagree</th>
<th>5 Slightly Agree</th>
<th>6 Agree</th>
<th>7 Strongly Agree</th>
<th>Mean</th>
<th>Std. Dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>SI1: People who influence my behaviour think that I should use technology.</td>
<td>0 (0%)</td>
<td>3 (4.4%)</td>
<td>14 (20.6%)</td>
<td>21 (30.9%)</td>
<td>18 (26.5%)</td>
<td>8 (11.8%)</td>
<td>4 (5.9%)</td>
<td>4.38</td>
<td>1.23</td>
</tr>
<tr>
<td>SI2: People who are important to me think that I should use technology.</td>
<td>2 (2.9%)</td>
<td>5 (7.4%)</td>
<td>8 (11.8%)</td>
<td>21 (30.9%)</td>
<td>15 (22.1%)</td>
<td>13 (19.1%)</td>
<td>4 (5.9%)</td>
<td>4.43</td>
<td>1.43</td>
</tr>
<tr>
<td>SI3: The administration of this university has been helpful in the use of technology.</td>
<td>1 (1.5%)</td>
<td>2 (2.9%)</td>
<td>6 (8.8%)</td>
<td>26 (38.2%)</td>
<td>20 (29.4%)</td>
<td>10 (4.7%)</td>
<td>3 (4.4%)</td>
<td>4.53</td>
<td>1.17</td>
</tr>
<tr>
<td>SI4: The university has supported the use of technology.</td>
<td>1 (1.5%)</td>
<td>1 (1.5%)</td>
<td>4 (5.9%)</td>
<td>18 (26.5%)</td>
<td>19 (27.9%)</td>
<td>19 (27.9%)</td>
<td>6 (8.8%)</td>
<td>4.97</td>
<td>1.23</td>
</tr>
</tbody>
</table>

Table 7 is a summary of descriptive analysis for the facilitating conditions in undergraduates’ technology use. It suggests that the undergraduates agree they have the necessary resources and knowledge to use technology. However, the undergraduates’ perceptions are rather neutral toward the available assistance when they encounter technology use difficulties.

Table 7: Descriptive Statistics for Facilitating Conditions (FC) (n= 68)

<table>
<thead>
<tr>
<th>Questionnaire Item</th>
<th>1 Strongly Disagree</th>
<th>2 Disagree</th>
<th>3 Slightly Disagree</th>
<th>4 Neither Agree Or Disagree</th>
<th>5 Slightly Agree</th>
<th>6 Agree</th>
<th>7 Strongly Agree</th>
<th>Mean</th>
<th>Std. Dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>FC1: I have the resources necessary to use technology.</td>
<td>0 (0%)</td>
<td>3 (4.4%)</td>
<td>4 (5.9%)</td>
<td>12 (17.6%)</td>
<td>26 (38.2%)</td>
<td>15 (22.1%)</td>
<td>8 (11.8%)</td>
<td>5.03</td>
<td>1.23</td>
</tr>
<tr>
<td>FC2: I have the knowledge necessary to use technology.</td>
<td>0 (0%)</td>
<td>1 (1.5%)</td>
<td>8 (11.8%)</td>
<td>10 (14.7%)</td>
<td>17 (25.0%)</td>
<td>22 (32.4%)</td>
<td>10 (14.7%)</td>
<td>5.19</td>
<td>1.28</td>
</tr>
<tr>
<td>FC3: When I encounter difficulties in using technology,</td>
<td>0</td>
<td>0</td>
<td>13</td>
<td>18</td>
<td>11</td>
<td>14</td>
<td>7</td>
<td>4.54</td>
<td>1.46</td>
</tr>
</tbody>
</table>
a specific person is available to provide assistance.

<table>
<thead>
<tr>
<th></th>
<th>0%</th>
<th>0%</th>
<th>19.1%</th>
<th>26.5%</th>
<th>16.2%</th>
<th>20.6%</th>
<th>10.3%</th>
</tr>
</thead>
</table>

FC4: When I encounter difficulties in using technology, I know where to seek assistance.

<table>
<thead>
<tr>
<th></th>
<th>0%</th>
<th>4%</th>
<th>6%</th>
<th>16%</th>
<th>17%</th>
<th>17%</th>
<th>8%</th>
<th>4.9%</th>
<th>1.36</th>
</tr>
</thead>
</table>

FC5: When I encounter difficulties in using technology, I am given immediate assistance.

<table>
<thead>
<tr>
<th></th>
<th>1%</th>
<th>10%</th>
<th>12%</th>
<th>21%</th>
<th>15%</th>
<th>5%</th>
<th>4%</th>
<th>4.03</th>
<th>1.40</th>
</tr>
</thead>
</table>

The undergraduates’ behavioural intention in technology adoption is statistically described in Table 8. As demonstrated, the undergraduates will use technology in the future. Moreover, they also agree that they have positive intention to use technology often in future or in the next few months.

Table 8: Descriptive Statistics for Behavioural Intention (BI) (n= 68)

<table>
<thead>
<tr>
<th>Questionnaire Item</th>
<th>1 Strongly Disagree</th>
<th>2 Disagree</th>
<th>3 Slightly Disagree</th>
<th>4 Neither Agree Or Disagree</th>
<th>5 Slightly Agree</th>
<th>6 Agree</th>
<th>7 Strongly Agree</th>
<th>Mean</th>
<th>Std. Dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>BI1: I intend to use technology in the next few months.</td>
<td>0 (0%)</td>
<td>1 (1.5%)</td>
<td>5 (7.4%)</td>
<td>9 (13.2%)</td>
<td>15 (22.1%)</td>
<td>14 (20.6%)</td>
<td>24 (35.3%)</td>
<td>5.59</td>
<td>1.36</td>
</tr>
<tr>
<td>BI2: I predict I would use technology in the next few months.</td>
<td>0 (0%)</td>
<td>2 (2.9%)</td>
<td>3 (4.4%)</td>
<td>7 (10.3%)</td>
<td>19 (27.9%)</td>
<td>13 (19.1%)</td>
<td>24 (35.3%)</td>
<td>5.62</td>
<td>1.34</td>
</tr>
<tr>
<td>BI3: I plan to use technology in the next few months.</td>
<td>0 (0%)</td>
<td>1 (1.5%)</td>
<td>5 (7.4%)</td>
<td>8 (11.8%)</td>
<td>12 (17.6%)</td>
<td>19 (27.9%)</td>
<td>23 (33.8%)</td>
<td>5.65</td>
<td>1.34</td>
</tr>
<tr>
<td>BI4: I will use technology in the future.</td>
<td>0 (0%)</td>
<td>1 (1.5%)</td>
<td>0 (0%)</td>
<td>5 (7.4%)</td>
<td>11 (16.2%)</td>
<td>20 (29.4%)</td>
<td>31 (45.6%)</td>
<td>6.09</td>
<td>1.08</td>
</tr>
<tr>
<td>BI5: I plan to use technology often.</td>
<td>0 (0%)</td>
<td>2 (2.9%)</td>
<td>4 (5.9%)</td>
<td>6 (8.8%)</td>
<td>15 (22.1%)</td>
<td>17 (25.0%)</td>
<td>24 (35.3%)</td>
<td>5.66</td>
<td>1.36</td>
</tr>
</tbody>
</table>

The descriptive statistics in Table 9 entails the undergraduates’ use behaviour. The analysis suggests that the undergraduates use technology for leisure, studies and daily communication. Interestingly, according to the statistics shown, the undergraduates’ technology use for course-related work overrides the use for other purposes.
Moreover, Table 10 is the summary of descriptive analysis for the UTAUT model. As shown in the table, use behaviour and behavioural intention scored the highest mean value (M = 5.72) followed by performance expectancy (M = 5.37, SD = 1.00) and effort expectancy (M = 5.13, SD = 1.09). Meanwhile, social influence indicated the lowest mean value (M = 4.58, SD = 1.00) whereas facilitating conditions demonstrated the second lowest mean value (M = 4.74, SD = 1.12). This shows that most of the students either use or have the intention to use the technology for course-related work, relaxation, and communication regardless of their social influence and facilitating conditions.

Table 10: Descriptive Statistics for UTAUT Model (n= 68)

<table>
<thead>
<tr>
<th>Variables</th>
<th>Mean</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Performance Expectancy</td>
<td>5.37</td>
<td>1.00</td>
</tr>
<tr>
<td>Effort Expectancy</td>
<td>5.13</td>
<td>1.09</td>
</tr>
<tr>
<td>Social Influence</td>
<td>4.58</td>
<td>1.00</td>
</tr>
<tr>
<td>Facilitating Conditions</td>
<td>4.74</td>
<td>1.12</td>
</tr>
<tr>
<td>Behavioural Intention</td>
<td>5.72</td>
<td>1.20</td>
</tr>
<tr>
<td>Use Behaviour</td>
<td>5.72</td>
<td>.81</td>
</tr>
</tbody>
</table>

Findings of this study also reported that there is a strong positive correlation between performance expectancy (r = .695, p < .0005), effort expectancy (r = .635, p < .0005), social influence (r = .544, p < .0005) and behavioural intention to use technology. In addition, there is also a positive correlation between facilitating conditions (r = .538, p < .0005) and use behaviour; and medium positive correlation between behavioural intention (r = .496, p < .0005) and use behaviour. Thus, the credibility of the UTAUT model in investigating the antecedents that influence technology use among Arts and Science undergraduate students is continuously being proven (Venkatesh & Zhang, 2010). This is also consistent with the research done by Venkatesh et al. (2003) a decade ago.
Lastly, an independent-samples t-test was carried out to compare the technology use between the Arts and Science undergraduate students. There was no significant difference ($t (66) = .558, p = .579$) found in terms of technology use for Arts ($M = 5.772, SD = .653$) and Science ($M = 5.661, SD = .980$) students which found to be concurrent with Williamson’s study (2011). The magnitude of the differences in the mean values (mean difference = .111) was very small (eta squared = .005).

6. Conclusion

Integrating technology in teaching and learning is to some extent an expectation in tertiary education. Technology is no longer regarded as novelty but a standard feature in the delivery of a course in tertiary institutions in Malaysia. However, there is a need to understand and identify the antecedents of technology use among graduates to help educators and education managers address the challenges and concerns experienced by them. However, this study was conducted with a modest sample size from two faculties within a university in Malaysia. Therefore it is not representative of the scenario in Malaysia. However, future research could explore the possibility of expanding the sample size or comparing Arts and Science undergraduates from different universities located in different countries.

Acknowledgements

We would like to thank Universiti Tunku Abdul Rahman, Malaysia for providing us the fund to conduct this project. This paper is also a part of the abstract entitled “Predictors of Behavioral Intention to Use Technology among Undergraduate Students” submitted and presented in the 2nd International Conference on Behavioral & Social Science Research 2014 under the same funding.

References


Liu, J. T., & Kostiwa, K. (2007). An application of the utaut model for understanding students perceptions using course management software. Communications of the IIMA, 7(2).


Equipping High School Students with the Abilities of Evaluating Evidence and Formulating Evidence for an On-line Decision-making Task

Shu-Sheng LIN\textsuperscript{a*}, Ying-Shao HSU\textsuperscript{b}
\textsuperscript{a}Graduate Institute of Mathematics and Science Education, National Chiayi University, Taiwan
\textsuperscript{b}Graduate Institute of Science Education, National Taiwan Normal University, Taiwan
\textsuperscript{*}lin-s-s@mail.nctu.edu.tw

Abstract: Preparing students with appropriate abilities for taking some on-line tasks is necessary. It will help students to build self-confidence and increase the probability to reach the success of learning. The purpose of this study aimed at equipping high school students with the skills of evaluating evidence and formulating evidence for taking an on-line task, in which students had to make a decision on choosing a location to build a reservoir. The developed instructional activity was a 4-hour unit, which provided a socioscientific context for students to understand the concept about evidence and discuss the reliability and validity of evidence to support if the global warming has been accelerating. The participant consisted of one earth science teacher and forty students. Two questionnaires and individual interviews were used to collect data. The results showed that the students had significant improvement in evaluating evidence, formulating evidence and justifying their arguments. Their understandings of criteria used for evaluating evidence became more clearly after the teaching. Most students reported that the instruction was beneficial for them to complete the decision-making task.

Keywords: Making decision, evidence evaluation, evidence formulation

1. Introduction

If a decision maker wants to make an evidence-based decision, he/she have to find evidence, evaluate evidence, use evidence, and justify his or her decision with evidence. These abilities are important but have been paid little attention in science curriculum and instruction. Therefore, Gott and Duggan (2007) advocated that science education has to get procedural and declarative understanding of evidence involved in science instruction. Prior studies have revealed that many of the students, from elementary students to university students, have difficulties in using evidence to support their arguments (Sandoval, Sodian, Koerber, & Wong, 2014), coordinating the claim or conclusion with the evidence (Zimmerman, 2007), and evaluating evidence (Nicolaidou, 2011). These abilities regarded as higher cognitive thinking skills are vital for students to deal with some issues, especially they are asked to make a decision in a socioscientific context. The socioscientific issue is one kind of scientific and social issues, which embed with some problems, controversies and dilemma caused by the application and development of science and technology (Zeidler et al, 2005), such as genetically modified organisms, radiation of mobile phones, or building a reservoir etc. People argue the solutions of the problem in an issue without reaching a conclusion. Zeidler and Nichols (2009) suggested that science teachers can select an appropriate socioscientific issue (SSI) for instruction, lead students to examine and discuss the arguments the stakeholders have, evaluate evidence of each argument with criteria, practice making a decision and justifying their decision with evidence. It means before making a decision on a SSI, there are many steps regarding understanding of and using of evidence involved in the process that are difficult for most of students. Hug and McNeil (2008), Schalk, Van der Schee and Boersma (2013) suggested that it is helpful for students to make a deliberate decision after they experience the instruction planned to improve their understanding of
evidence and skills of evaluating and using evidence. In this study, we attempt to equip high school students with the skills of evaluating evidence and formulating evidence before they take an on-line decision-making task.

2. Method

2.1 Participants

The participants consisted of one experienced teacher and forty grade eleven students (27 girls and 13 boys). The earth science teacher has 20-year teaching experiences. She joined the workshop held by the researcher to learn the declarative and procedural knowledge about evidence, the operation and the contents of software, and to discuss the teaching materials and methods with the researcher. The students whose age was 16-17 years old did not have the formal experiences to evaluate evidence, formulate evidence and justify arguments with evidence before.

2.2 The Instruction

The instruction was to enhance the students’ understanding of the concept of evidence, and improve their abilities to evaluate evidence, formulate evidence and justify arguments with evidence. The instructional unit included four hours. At the first two hours, the teacher led the students to discuss the importance of evaluating and using evidence in everyday life. After the criteria used to evaluate the reliability and validity of evidence were developed through group discussion, the students assessed the criteria each group formulated for their appropriateness through the whole class discussion. During the last two hour, one text provided for the students to read includes five stakeholders and their arguments to the question – “Global warming becomes more serious than before. Is it man-made? Each argument had at least one piece of evidence to support that it is caused by man-made. The students were led to discuss the reliability of evidence, the relevance between the evidence and the claim, and how to make much stronger evidence to support or rebut the claim and justify their arguments.

2.3 The software and the task

The students were asked in a software environment to complete an evidence-based decision-making task, in which they had to choose an appropriate location to build a reservoir within a limited time. During the trial-and-error process of making decision, they had to use the abilities of formulating criteria, selecting and evaluating evidence, and justify their choice with evidence. Figure 1 shows one of the interfaces of “Constructing Reservoir” software.

![Fig.1 The interface of “Constructing Reservoir” software](image)

2.4 Instruments

The questionnaire “Evidence evaluation and use” included a scientific research context and three
questions. The context described three animal studies on exploring the function of onions, the results and the consistent conclusions. Three questions were used to assess the students’ abilities to generate an argument, evaluate the reliability of the evidence, and formulate evidence to support the arguments. The questions are: (1) Do you agree or disagree with the conclusion of three animal studies? Why? (2) Do you think the evidence the author described is reliable to support the conclusion? Explain why in detail. (3) If the scientists can make more evidence to support the conclusion, do you think what it is? Explain your reasons. The pre- and posttest administered to the students before and after the instructional intervention were the same. The other questionnaire is “Reflection on Learning” used to collect the students’ feedback about the instruction. It was related to their attitude towards the teaching contents, methods and their reflection on learning for completing the decision-making task.

3. The Results

A series of t-test were run to examine the improvement of the students’ abilities after instructional intervention (Table 1). The findings showed that the students had statistically significant improvement in scores for making warrants, evaluating the reliability of the evidence, and formulating evidence to support the arguments ($p<.01$).

<table>
<thead>
<tr>
<th>Questions</th>
<th>Pretest Mean(S.D.)</th>
<th>Posttest Mean(S.D.)</th>
<th>t value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Making warrants</td>
<td>1.40(0.81)</td>
<td>2.15(1.19)</td>
<td>4.39(0.00**)</td>
</tr>
<tr>
<td>Evaluating evidence</td>
<td>1.35(0.92)</td>
<td>2.13(0.97)</td>
<td>5.69(0.00**)</td>
</tr>
<tr>
<td>Formulating evidence</td>
<td>1.43(0.68)</td>
<td>2.17(0.96)</td>
<td>4.13(0.00**)</td>
</tr>
<tr>
<td>Total Scores</td>
<td>4.18(1.89)</td>
<td>6.45(2.09)</td>
<td>12.23(0.00**)</td>
</tr>
</tbody>
</table>

Figure 2 and 3 respectively revealed that the criteria the students used for evaluating evidence were a few different before and after the instruction. The criteria of “time” used in the posttest instead of in the pretest. The criteria the students used for formulating evidence were the same in the pretest and posttest. The number of criteria appeared in posttest had been significantly increased than in the pretest in both abilities.

Meanwhile, according to the feedbacks the students expressed in the questionnaire, most of the students pointed that they benefited a lot from the instruction. For example, the student S03 said that “if without the instruction before I took the decision-making task, I nearly did not have the idea to use criteria to evaluate and select evidence to help myself to complete it.”
4. Conclusions

The instructional design in this study supports the students to develop the abilities in making arguments, evaluating evidence and formulating evidence to support their arguments. Based on the students’ feedback and individual interviews, it is really helpful for the students to apply these abilities to take the decision-making task in the software environment.

However, it is not enough to proof that all of the students transfer these abilities well for taking the task in this pilot study. Therefore, the researcher further plans to adopt “two-group pretest-posttest experimental design”. The experimental group receives the instruction this study showed. The control group receives the instruction without emphasizing on learning to develop criteria for evidence evaluation. Through the comparison we can confirm the effect of instructional intervention. Moreover, the researchers will examine what strategies the students adopt during the period of taking the task. It will reveal the abilities the students apply in the task.

References


An Investigation of Relationships between Biology Attitudes and Perceptions toward Instructional Technology in Analogy-based Simulation on Light Reaction

Sarunya PINATUWONG* & Niwat SRISAWASDI*

*a Science and Technology Education Program, Faculty of Education, Khon Kaen University, Thailand
b Division of Science, Mathematics, and Technology Education, Faculty of Education, Khon Kaen University, Thailand

*niwsri@kku.ac.th

Abstract: Computer-simulated scientific phenomena have become an indispensable tool in modern scientific investigations and in contemporary science class. In context of science-based education, computer simulation or visualization are now commonly used to promote student's meaningful understanding in science concepts, and motivation to learn science. This study presents the development of an analogy-based simulation, as a novel pedagogy-oriented simulation, for biology learning of light reaction phenomenon. 46 eleventh-grade students were recruited to participate interacting with the analogy-based simulation. The 25-item Likert-scale questionnaire on attitude toward biology lesson scale was administered to the student before interacting the simulation, and another 21-item Likert-scale questionnaire on perception toward instructional technology was completed by the student after the interacting. Overall finding of this study suggests that students' attitudes towards biology did not determine their perceived learning, flow, enjoyment, perceived ease-of-use, perceived usefulness, and satisfaction, delivered by the analogy-based simulation. As such, this study concludes that the analogy-based simulation could be used in biology class disregarding students' attitudes toward biology. The findings could be implied for designing effective computer simulation to facilitate biology teaching and learning in school science.

Keywords: Computer simulation, analogy, perception, attitude, biology education

1. Introduction

Currently, the prospect of research on the use of ICT in education in general, or even in the specific case of science education is widespread, especially studies on the use of computer simulations. Nowadays, computer simulation has become increasingly powerful and available to teachers in the past decade (de Jong, Linn and Zacharia, 2013; Trundle and Bell, 2010). For instructional context, computer-simulated technology has been used to facilitate teaching and learning by visualizing objects, processes, and interact dynamics models of natural phenomenon, that are normally beyond the user’ control in the natural world (de Jong, Linn and Zacharia, 2013; Perkins et al., 2006; Wieman, Perkins and Adams, 2008). These technology offer idealized, dynamic and visual representations of invisible phenomena and experiments which would be dangerous, costly or otherwise not possible in school laboratories (Hennessy, 2007). Since, simplified versions of the natural world were showed by computer simulation, students’ attention can be more focus directly on the desired phenomenon (de Jong and van Joolingen, 1998; Perkins et al., 2006; Wieman, Perkins and Adams, 2008).

As such, successful concepts of simulation-based teaching and learning have been reported by means of discovery learning (de Jong, Linn and Zacharia, 2013) and inquiry-based learning (Perkins et al., 2006; Wieman, Perkins and Adams, 2008). The researchers have interested to science learning with analogy, which is an effective pedagogy to assist students learning about an unfamiliar concept, system, or process, called target by means of its relationship to a familiar concept, system, or process,
called analog. Based on literature reviewing, very few study (i.e. Ashe and Yaron, 2013; Ünlü and Dökme, 2011) has investigated effects of analogy-based simulation that may influence students’ conceptual development in science. Especially, considering to the nature of biology knowledge, lot of invisible processes and biological mechanism that occurred in organisms was presented to student and they encountered with learning difficulties about biology. Teaching biology through analogy is an instructional idea with aim to help student learn biology meaningfully. Schiff (1970) stated that perception and attitudes were related together in which attitudes affect perception, perception affects attitudes, and cognition plays a role in both of them. An important consequence of instruction is the student’s attitude toward the subject and the previous study found that there is usually positive correlation between attitudes and achievement, but researcher cannot assume a positive attitude on the basis of achievement alone (Russell and Hollander, 1975). It is possible for a student with low achievement to have developed a very positive attitude toward the subject matter, but it is also possible that a student who indicated on post-test that they have learned the subject matter well may also have learned to dislike or, worse yet, hate the content (Russell and Hollander, 1975). From these reasons, the researchers have created a computer simulation emphasized analogy approach on biology concept of light reaction and an evaluation of students’ perception towards the simulation regarded their existing biology attitudes.

2. Analogy-based Simulation

Analogy-based simulation is computer-simulated visualization that does not directly depict atoms or molecules; rather, they simulate a different physical system that relates to a chemical system or concept by analogy. For examples, Ashe and Yaron (2013) created computer simulation visualized chemistry phenomena by making analogies between chemical systems and familiar objects from students’ everyday experiences, e.g. boxes, steps, and balls. Because students are not typically experienced in reasoning qualitatively about chemical systems, the analogy-based simulation aims to leverage students’ experience with the world around them to help them better understanding of scientific concepts (Ünlü and Dökme, 2011). As the way it was created and students are interactive and dynamic by their nature, analogy-based simulations are likely to be more engaging to students than static, non-interactive presentations of analogies. Moreover, since students cannot benefit from an analogy if they do not engage with it, simulations may offer an advantage here. There are dual-situated events in which analogy could be used to incorporate into computer simulation. Firstly, analogies were used to simulate concepts which are not possible to simulate directly with portrayal of atoms or molecules, or which could be simulated more clearly by analogy. Another, analogy could be used to address students’ understanding for important knowledge abstractions (Ashe and Yaron, 2013).

3. Methods

3.1 Participants

A total of 46 student-respondents in their eleventh grade, age ranging from 16 to 17 years in a local public school at the Northeastern region of Thailand participated in this study. They have no experience yet using analogy-based simulation. This implied that they are heterogeneous before interacting with the simulation.

3.2 Learning Materials

Figure 1 illustrates examples of analogy-based simulation on light reaction phenomenon of plant photosynthesis. As aforementioned, target and analog are important concepts for analogy approach, and both were created into the computer simulation. In this analogy-based simulation, the target event was electron transfer during photosystem I (PS-I) of light reaction and the analog event was coal transfer on a pirate ship. In an addition, there were two parameters which student could vary for their investigation; wavelength of light and light intensity.
3.3 Procedures

Before exposing to the analogy-based simulation, the students took a 25-item five-point Likert-scale questionnaire on attitude toward biology lesson scale for 10 minutes. A 25-item Likert-scale questionnaire was developed to use in this study for examining students’ interest in biology lessons (IBL), understanding and learning biology (ULB), importance of biology in real-life (IBR), biology and occupational choice (BOC) toward biology learning. There were six items of IBL, 10 items of ULB, five items of IBR and four items of BOC (Ayyıldız and Tarhan, 2013). To develop a Thai version of the questionnaire, the original English version was translated identically in Thai language. One expert was recruited to identify communication validity of the items. On each item, respondents were assigned to rate how the respondent agree with into five scale, from 1-strongly disagree to 5-strongly agree. The reliability for Thai version on IBL, ULB, IBR, BOC was 0.79, 0.70, 0.70, 0.65 respectively, and the overall was 0.76. To experience with the analogy-based simulation, they were assigned to interact with the simulation independently for 30 minutes. After that, the students took 21-item five-point Likert-scale questionnaire measured perception toward instructional technology for 10 minutes. The questionnaire was developed to examine students’ perceived learning (PL), flow (FL), enjoyment (EJ), perceived ease of use (PEOU), perceived usefulness (PU), and satisfaction (ST). There were four items of PL, five items of FL, three items for EJ obtained from Barzilai and Blau (2014), and three item of PEOU, three items of PU, three items of ST obtained from Cheng (2014). The reliability for Thai version on PL, FL, EJ, PEOU, PU, ST was 0.80, 0.82, 0.75, 0.74, 0.84, 0.77 respectively and the overall was 0.95. To examine correlation between students’ attitude and perception toward instructional technology, Pearson’s correlation was performed in SPSS 21.0.

4. Results and Discussion

Table 2 shows Pearson's correlations among the variables. Regarding the Pearson’s correlation analysis of attitudes towards biology, the results indicated that correlation among PL, FL, EJ, PEOU, PU, and ST, reveals a significant positive correlation ($p$-value < 0.01), as well as the correlation among IBL, ULB, IBR, and BOC in perceptions towards instructional technology ($p$-value < 0.01). However, there was no significant correlation between variables of attitudes towards biology and variables of perceptions towards instructional technology, except the product-moment correlation between perceived ease of use (PU) and interest in biology lessons (IBL), $r=0.294$, $p$-value < 0.05. For the overall summarization, these findings suggest that students’ attitudes towards biology did not determine their perceptions towards instructional technology delivered by the analogy-based simulation. This means the use of analogy-based simulation for biology learning could support students’ perceptions even individual student who have a positive or negative biology attitudes. This evidence consistent with previous research finding that students perceived learning with simulation offer instructional advantages and they can gain benefits for their experiences (Baillie and Curzio, 2009; Prokop, Prokop and Tunnicliffe, 2007). Furthermore, the use of instructional technology could be a challenge in teaching methods for providing interactive classroom activities and stimulate students’ engagements and attitudes toward science (Hansen and Birol, 2010).

Table 1: Descriptive and Pearson product-moment correlation matrix of attitudes towards biology lesson and perceptions towards instructional technology.
5. Conclusion

To facilitate students’ learning in biology, an analogy-based simulation was created to assist students learning about an unfamiliar concept, system, or process. The result indicated that the use of analogy-based simulation for biology learning to students could be benefits on their perceptions even holding a positive and negative attitude towards biology lessons. Thus, it is crucial to design computer-simulated visualization by the use of analogy approach to scaffold students’ perceptions. To this end, researchers should pay more attention on principle of analogy when designing computer simulation to enrich the learning content for biology learning in school science.

Acknowledgements

This work was financially supported by Institute for the Promotion of Teaching Science and Technology (IPST) and Graduate School, Khon Kaen University. The author would like to express gratefully acknowledge to Faculty of Education, Khon Kean University, for supporting this contribution.

References


Visual Behavior and Cognitive Load in Augmented Reality Learning Environment

Wei-Yan LIN *, Meng-Jung TSAI & An-Hsuan WU
Graduate Institute of Digital Learning and Education
National Taiwan University of Science and Technology, Taiwan
*M10211004@mail.ntust.edu.tw

Abstract: This pilot study was attempted to explore the relationships among cognitive load, visual behavior, and reading comprehension in augmented reality (AR) learning environment. An ASL MobileEye-XG eye-tracking system was used to record the participants’ visual behaviors while they read Taiwan traditional folk belief information demonstrated by AR technologies. A survey was used to examine participants’ cognitive loads due to different cognitive elements (videos, texts and pictures) designed in the AR contents. Currently, the results showed that participants with higher reading comprehensions seemed to have higher cognitive loads from videos. Although no significant relationships were found between visual behaviors and reading comprehensions, some significant correlations were found between visual behaviors and cognitive loads, and also between cognitive loads and reading comprehensions. More results will be discussed in the conference.

Keywords: Cognitive load, visual behavior, reading comprehension, augmented reality

1. Introduction

Augmented reality (AR) technologies, which utilize image recognition technology to link computer-generated virtual elements with real-world elements and display two types of elements on the same platform at the same time, are gradually promoted to be used in educational settings (Chen & Tsai, 2012; Wu, Lee, Chang, & Liang, 2013) in different learning domains. This study was attempted to use augmented reality technology in Taiwanese folk belief exhibition to explore its possibility of use in multicultural education, especially for the cultural inheritance of Taiwan traditional folk beliefs.

Eye movement studies have been applied to learning for decades and have caused increasing attentions in educational research recently (Lai et al., 2013). Visual attentions recorded by eye-tracking systems have been found to be significantly related to learning performance in environments where pictures and texts were displayed simultaneously (Yang, Chang, Chien, Chien, & Tseng, 2013; Ho, Tsai, Wang, & Tsai, 2014). However, little studies explored the situations that videos were added simultaneously as used in AR learning environments. Thus, this study added video into learning environment along with picture and text. Besides, eye tracking measures seemed to be found relating to cognitive load, but didn’t analyze which type of cognitive load related to participants’ eye movements (Wang, Yang, Liu, Cao, & Ma, 2014). According to the literature, cognitive load can be classified into three different types, i.e., intrinsic cognitive load, extraneous cognitive load, and germane cognitive load (Sweller, van Merriënboer, & Paas, 1998). Therefore, this pilot study aimed to explore the relationships among different types of cognitive loads, visual behaviors, and reading comprehensions in a Taiwanese folk belief exhibition which includes texts, pictures, and videos linked by AR.

2. Method

2.1 Experiment

Seven university students, four males and three females, in north Taiwan were volunteered to participate in a pilot eye-tracking experiment. The average age of the participants was 23.71(SD =
All of them had little knowledge regarding Taiwanese traditional folk beliefs and little experience in participating related traditional cultural events. Therefore, the participants had about the same level of prior knowledge or experience regarding the learning contents.

The learning material of this study was set as a real Taiwanese folk belief exhibition. There were eight ceremonies’ introductions shown as posters in this exhibition. Every ceremony was introduced by one picture with a short caption below, one paragraph of text aside the picture that describes how the ceremony held, and one video showing the procedure of real ceremony situation linked by the augmented reality technology. The videos were designed to be activated by markers on posters and shown in a smart phone.

2.2 Instrument

In order to investigate different types of cognitive load from different types of materials (videos, texts, pictures), nine questions were designed to examine the types of cognitive loads. A self-reported 0-10 scale was used to evaluate each of the questions. The larger the number, the higher the cognitive load was perceived. Besides, an ASL MobileEye-XG system (60 Hz) was used to record the participants’ eye movement data during the participants viewing the AR exhibition. After the experiment, a video analysis was used to analyze each participant’s visual behavior data. Total reading time on each type of learning materials, total entered count on each type of learning materials, and average entered duration of each type of learning materials were calculated based on the scan path video of each participant. There were total twelve visual behavior measures coded in current analysis. Finally, the posttest in this study was assessed by twenty multiple-choice questions concerning the concepts demonstrated in the AR exhibition.

2.3 Data Collection and Analyses

The procedure of this study’s experiment can be divided into four parts. First, each of the participants was taught how to link the video by AR with a smart phone. Second, an eye-tracking calibration was then conducted. Third, the participant started to view the learning materials for twenty minutes. At last, the participant rated the cognitive load questionnaire and received the posttest with a comprehension test of twenty questions. As for data analyses, Mann-Whitney U test was used to analyze the differences of cognitive loads and visual behaviors between high performance group and low performance group. Furthermore, Spearman’s correlation analyses were used to analyze the relationships among the participants’ cognitive loads, visual behaviors, and reading comprehension.

3. Result

According to the result of Mann-Whitney U test, it was found that there was significant difference between high performance group and low performance group toward extraneous cognitive load from video (z = -1.999, p = .046). The participants in high performance group had higher extraneous cognitive load from video (Mean = 5.250, SD = 2.062) than the participants in low performance group (Mean = 1.667, SD = 1.155). Moreover, it was found that there was no significant difference between high performance group and low performance group on visual behaviors.

In addition, regarding the result of Spearman’s correlations analysis between the participants’ cognitive load and visual behaviors, the participants’ extraneous cognitive load from picture had significantly negative correlation with total entered count of picture (r = -.805, p < .05). This finding reveals that the participants who had higher extraneous cognitive load from picture, they had lower counts of looking pictures. It was also found that the participants’ extraneous cognitive load from text had significantly positive correlation with average entered duration of pictures’ short explanation. This finding reveals that the participants who had higher extraneous cognitive load from text, they had longer average entered durations on the short explanations of pictures.

Finally, based on the result of Spearman’s correlations analysis between the participants’ visual behaviors and reading comprehension, it was found that there was no significant relationship between the participants’ visual behaviors and reading comprehension. However, according to the result of
Spearman’s correlations analysis between the participants’ cognitive load and reading comprehension, it was found that the participants’ extraneous cognitive load from video had significantly positive correlation with the participants’ reading comprehension ($r = .972, p < .001$). This finding reveals that the participants had higher extraneous cognitive load from video might get higher reading comprehension.

4. Conclusion and discussion

According to the results, the participants in high performance group had higher extraneous cognitive load from video than the participants in low performance group. In addition, the participants had higher extraneous cognitive load from video might get higher reading comprehension. In opposition to general viewpoint, which suggests that extraneous cognitive load should be reduced for better learning (Sweller & Chandler, 1994), it seemed that extraneous cognitive load might actually help the participants to learn better.

In addition, the results showed that the participants who had higher extraneous cognitive load from picture had lower counts of looking pictures. It seems to mean that learners who think that displaying learning contents with pictures is not good for learning might not take pictures as main learning materials; on the contrary, learners who think that displaying learning contents with pictures is good for learning might take pictures as helpful learning materials. Furthermore, the participants who had higher extraneous cognitive load from text tended to have longer average entered durations on the short captions of pictures. It indicates that learners who think that displaying learning contents with texts is not good for learning might need to take more time on realizing what short captions of pictures are talking about; on the contrary, learners who think that displaying learning contents with texts is good for learning might need to take less time on realizing what captions of pictures are talking about. It is interesting to notice that extraneous cognitive load might be the main element that influences the participants’ viewing behaviors in AR learning environments.

Furthermore, opposite to the previous study (Yang et al., 2013), it was found that there was no significant relationship between the participants’ visual behaviors and reading comprehension in this study. It might be caused by the small number of participants that made the statistical number too hard to be significant. It is suggested that the relationship between visual behaviors and reading comprehension should be paid more attention. More details of the results will be presented and discussed in the conference.

References

Visual Behavior and Cognitive Load on E-book Vocabulary Learning

An-Hsuan WU*, Po-Fen HSU, Hui-Jou CHIU & Meng-Jung TSAI
Graduate Institute of Digital Learning and Education, National Taiwan University of Science and Technology, Taiwan
*M10111022@mail.ntust.edu.tw

Abstract: This pilot study examined learning English vocabulary by an E-book from cognitive load aspect, and adopted eye-tracking data as a reference to investigate participants’ attention distribution among learning targets, pictures and background graphics. An ASL Mobile Eye-XG eye-tracking system was used to record participants’ visual behaviors in this study. A Wilcoxon test was conducted to compare the performance before and after the learning process. Correlation tests were conducted to examine the relationships among posttests, visual behaviors and cognitive loads. Also an ANOVA analysis was conducted to compare the total reading time spent on texts, pictures and background graphics. The preliminary results indicated that participants demonstrated a significant progression through reading the E-book. Besides, it was found that the participants spent more time on reading texts than pictures and background graphics. More details of the results will be discussed in the conference.

Keywords: E-book, cognitive load, eye-tracking, English vocabulary learning, visual behavior.

1. Introduction

Owing to the convenience and motility, the rapid growth of mobile devices generated. Accompany with this state, it springs up varied kinds of products that fit into the mobility mobile devices. E-Book is one of them, and plays an important role in it. Moreover, it pushes the relative studies forward. Mayer’s Multimedia Learning Theory became a reference for multimedia learning materials design and examining in many studies. Mayer’s (2009) Spatial Contiguity Principle and Temporal Contiguity Principle (2009) claimed that to present integrated texts and graphics on the same page at the same time will be a support during cognitive processing. Furthermore, it may lead to a better learning performance since learners have distributed their cognitive resource to deal with connecting the relation between texts and graphics rather than searching the position of target texts and the corresponding graphics. In addition, the Segmenting Principle, which was also proposed by Mayer, noticed that the key point to carry out this principle is to design a self-paced learning material. On the other side, the cognitive load theory proposed by Sweller, van Merrienboer, and Pass (1998) distinguished cognitive load into intrinsic cognitive load, extraneous cognitive load, and Germane cognitive load (Sweller, et al., 1998; Paas, Tuovinen, Tabbers, and Van Gerven, 2003). Further, van Gog & Scheiter (2010) hold the view that eye tracking could be a direct measurement to help explain the cognitive load during learning process. According to the reviewed paper of Lai et al (2013), studies were prompt in adopting eye tracking from 2009, and also mentioned that many studies suggested using eye-tracking measure as index to combine with interpreting cognitive or meta-cognitive concerned learning. Overall, this study aims to inquire into the relation between eye-tracking measure and cognitive load throughout learning English vocabulary by E-book and also examine the interaction among texts, pictures, and background graphics (hereinafter referred to as background) while reading an English vocabulary E-book.
2. Method

2.1 Participants

Six participants (three males and three females, aged twenty and over) from an university of north Taiwan got involved in this study. Their English ability was under the standardized test – GETP intermediate level (equal to CEFR B1 threshold).

2.2 Instrument

An eighteen English stationary and kitchenware vocabularies E-book is the leaning material of this study. All the participants read the same eighteen vocabularies, yet these vocabulary cards’ orders were randomized for individuals. In addition, this study adopted ASL Mobile Eye-XG eye-tracking system, with a sapling rate of 60 Hz, which recorded sixty eye movements information per second. It allowed participants to read the E-book comfortable that close to real reading state without limiting their movement. Meanwhile, the difficulty of each vocabulary that the participants perceived when they were answering the pretest sheet was also asked following by the meaning of each vocabulary. And participants rated a self-reported number lines cognitive load questionnaire after the experiment. The cognitive questionnaire consisting of nine questions, participants rated from one to ten, which indicated that, the vocabulary difficulty from very easy to very difficult, and the degree of assistance and interference about pictures and backgrounds from very low to very high. Finally, the pretest sheet contains all the vocabularies of E-book. Participants were asked to fill in the blank with corresponding words of vocabulary in Chinese and rated the difficulty of each vocabulary at the same time. For the posttest, they were also asked to fill in the blank with corresponding Chinese words, but the order was different from the pretest sheet by randomizing.

2.3 Data Collection and Analyses

Participants were asked to fill in pretest sheet at first. Then, they went through on 5-point eye calibration. After calibration, they began to read the English vocabulary E-book with a time limit of ten minutes. Last, they were asked to fill out a posttest sheet and a cognitive load questionnaire. As for the data analyses, this study conducted Wilcoxon test to compare the pretest and posttest. Besides, correlation analysis was used to examine the relationship between visual behavior and cognitive load, and visual behavior and posttest. In addition, an ANOVA analysis was conducted to compare the total reading time and total entered count of background, pictures, and texts. Besides, recording videos of vocabularies learning behavior were analyzed using Interact 9 software. Statistics of total reading time, entered count and duration of backgrounds, texts, and pictures were coded.

3. Result

3.1 Wilcoxon test of pretest and posttest

As shown in Table 1, through a Wilcoxon test, pretest (M=50.50, SD= 10.436) has significant different from posttest (M=97.71, SD= 2.704) (z=-2.207 , p=0.027, p <0.05) , and further comparing the mean score of each, we can learn that the posttest is higher than pretest. That is, it has enhanced the English vocabulary learning performance through the designed E-book in this study.

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>Mean difference</th>
<th>SD</th>
<th>z</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>pretest-posttest</td>
<td>6</td>
<td>-47.208</td>
<td>9.592</td>
<td>-2.207</td>
<td>0.027</td>
</tr>
</tbody>
</table>

* p < 0.1, ** p < 0.05, *** p < 0.01
3.2 Correlation between Posttest, Visual behavior, and Cognitive load.

As shown in Table 2, Background Assistance had significantly negative correlations with Posttest ($r=-.939 \cdot p=.005 \cdot p<.05$). This finding reveals that participants who had better performance on learning English vocabulary perceived less assistance from backgrounds. On the contrary, participants who had worse performance perceived more assistance from backgrounds.

Table 2: Correlation between Posttest and Cognitive load

<table>
<thead>
<tr>
<th></th>
<th>PiA</th>
<th>PrA</th>
<th>BA</th>
<th>D</th>
<th>PiI</th>
<th>PrI</th>
<th>BI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Posttest</td>
<td>-0.485</td>
<td>-0.127</td>
<td>-0.939**</td>
<td>0.000</td>
<td>-0.211</td>
<td>0.018</td>
<td>-0.949**</td>
</tr>
</tbody>
</table>

* $p < 0.1$, ** $p < 0.05$

Note: PiA=Picture Assistance, PrA= Pronunciations Assistance, BA=Background Assistance, D=Difficulty of vocabularies, PiI= Pictures Interference, PrI= Pronunciation Interference, BI=Background Interference

3.3 Correlation between Visual behavior and Cognitive load

As shown in Table 3, Total Entered Count of Background had significantly negative correlations with Pronunciation Interference ($r=-.845 \cdot p=.034 \cdot p<.05$). This finding reveals that during the reading process, with paying more attention to backgrounds, participants perceived less pronunciation interference, and with paying less attention to backgrounds, participants perceived more pronunciation interference relatively.

Table 3: Correlation between Visual behavior and Cognitive load

<table>
<thead>
<tr>
<th></th>
<th>PiA</th>
<th>PrA</th>
<th>BA</th>
<th>D</th>
<th>PiI</th>
<th>PrI</th>
<th>BI</th>
</tr>
</thead>
<tbody>
<tr>
<td>TEC of Background</td>
<td>0.696</td>
<td>-0.441</td>
<td>-0.232</td>
<td>0.696</td>
<td>-0.683</td>
<td>-0.845**</td>
<td>0.098</td>
</tr>
<tr>
<td>TET of Background</td>
<td>-0.116</td>
<td>0.383</td>
<td>-0.348</td>
<td>-0.058</td>
<td>-0.293</td>
<td>-0.845**</td>
<td>0.098</td>
</tr>
<tr>
<td>TEC of Text</td>
<td>-0.116</td>
<td>0.647</td>
<td>0.145</td>
<td>-0.493</td>
<td>-0.488</td>
<td>-0.507</td>
<td>0.293</td>
</tr>
<tr>
<td>TET of Text</td>
<td>-0.319</td>
<td>0.765*</td>
<td>-0.029</td>
<td>-0.493</td>
<td>-0.293</td>
<td>-0.676</td>
<td>0.293</td>
</tr>
<tr>
<td>TEC of Picture</td>
<td>-0.162</td>
<td>0.493</td>
<td>-0.338</td>
<td>-0.441</td>
<td>-0.396</td>
<td>-0.429</td>
<td>-0.198</td>
</tr>
<tr>
<td>TET of Picture</td>
<td>0.145</td>
<td>0.088</td>
<td>-0.464</td>
<td>-0.029</td>
<td>-0.488</td>
<td>-0.507</td>
<td>-0.293</td>
</tr>
</tbody>
</table>

* $p < 0.1$, ** $p < 0.05$

Note: PiA= Pictures Assistance, PrA= Pronunciations Assistance, BA=Background Assistance, D=Difficulty, PiI= Pictures Interference, PrI= Pronunciation Interference, BI=Background Interference

TEC of Background= Total Entered Count of Background, TET of Background= Total Entered Time of Background, TEC of Text= Total Entered Count of text, TET of Text= Total Entered Time of Text, TEC of Picture= Total Entered Count of Picture, TET of Picture= Total Entered Time of Picture

3.4 Correlation between Visual behavior and posttest

As shown in Table 4, there are no significant correlation among posttest and eye-tracking measures.

Table 4: Correlation between Visual behavior and posttest

<table>
<thead>
<tr>
<th></th>
<th>TRT</th>
<th>TRP</th>
<th>TEC of Background</th>
<th>TET of Background</th>
<th>TEC of Text</th>
<th>TET of Text</th>
<th>TEC of Picture</th>
<th>TET of Picture</th>
</tr>
</thead>
<tbody>
<tr>
<td>Posttest</td>
<td>.031</td>
<td>.309</td>
<td>-.062</td>
<td>.123</td>
<td>-.370</td>
<td>-.185</td>
<td>.172</td>
<td>.278</td>
</tr>
</tbody>
</table>

* $p < 0.1$, ** $p < 0.05$

TRT= Total Reading Time, TRP= Total Reading Pages, TEC of Background= Total Entered Count of Background, TET of Background= Total Entered Time of Background, TEC of Text= Total Entered Count of text, TET of Text= Total Entered Time of Text, TEC of Picture= Total Entered Count of Picture, TET of Picture= Total Entered Time of Picture
3.5 ANOVA results

In order to compare the total reading time and total entered count on backgrounds, texts, graphics, a factorial ANOVA was conducted. The results indicate that the all participants spent longer reading texts than looking at backgrounds. On the other hand, the frequency of reading texts and looking at graphics are both more than looking at backgrounds.

Table 5: Total Reading Time and Total Entered Count on background, text and graphic

<table>
<thead>
<tr>
<th></th>
<th>(1)Background</th>
<th>(2)Text</th>
<th>(3)Graphic</th>
<th>F</th>
<th>Scheffe</th>
</tr>
</thead>
<tbody>
<tr>
<td>index</td>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
</tr>
<tr>
<td>TRT</td>
<td>22.512</td>
<td>9.347</td>
<td>81.870</td>
<td>55.067</td>
<td>66.845</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TEC</td>
<td>47.000</td>
<td>17.251</td>
<td>120.167</td>
<td>46.154</td>
<td>121.167</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* p < 0.1, ** p < 0.05

Note: TRT= Total Reading Time, TEC= Total Entered Count

4. Conclusion and discussion

This study aims to investigate the correlation among visual behavior, cognitive load, and performance during the process of reading an English vocabulary E-book. In sum, according to the results, it is an effective English vocabulary E-book to help learners to learn English vocabulary by themselves. As for cognitive of background shows negative correlation with posttest, which means that to some extent, it seems that background has interaction with learners’ cognitive load of background and performance. The further study can probe the interaction between the background design and the performance and relative reading behavior.

In another aspect, the significant negative correlation among frequency and duration looking at backgrounds with the cognitive load of pronunciation interference and the total reading time of texts shows significant positive correlation with pronunciation assistance that the participants perceived. For this result, we speculate that participants thought that pronunciation function was helpful while they integrated elements, which are more relative to each other (texts and pronunciations). However, while integrated elements that are less relative to each other (backgrounds and pronunciations) they would take it as interference. Last, with the result of total reading time and total entered count on reading texts and looking at pictures and backgrounds, we can know that learners would primly focus on texts and pictures, while background plays a subordinate role.

Further study can be conducted to involve more participants to infer to larger population. On the other hand, the collected eye-tracking data can be further interpreted by deeper analysis. More details of results will be presented and discussed in the conference if the paper is accepted by the conference.

References


Incorporating augmented reality into learning practical skills for medical surgery

Ying-Shao Hsu\textsuperscript{a}, Yuan-Hsiang Lin\textsuperscript{b}, Beender Yang\textsuperscript{c}, Shih-Fan\textsuperscript{b} Yang, Ya-Yen Chan\textsuperscript{b}, Zi-Hao Lin\textsuperscript{b}, and Yi-Sheng Chan\textsuperscript{b}

\textsuperscript{a}Graduate Institute of Science Education, National Taiwan Normal University, Taiwan
\textsuperscript{b}Department of Electronic and Computer Engineering, National Taiwan University of Science and Technology, Taiwan
\textsuperscript{c}Saturn Imaging Inc., Taiwan

*yshsu@ntnu.edu.tw

Abstract: In this paper, we demonstrate how to incorporate augmented reality (AR) into the learning of practical skills for medical surgery. We embedded AR in authentic inquiry activities so that students could experience when and how to carry out a certain medical surgical procedure. Two learning modules related to medical surgery were developed, "laparoscopic surgery" and "cardiac catheterization". Thirty-two senior high school students participated in this study and their perceptions of AR were examined. A survey of student perceptions included the three constructs of authenticity, engagement and motivation. The results showed that the students had positive perceptions (overall mean = 4.1) of AR after completing the two modules. However, AR authenticity was the concept perceived as having the lowest ranking (mean = 3.7). In contrast, both the motivation triggered by AR and engagement reached 4.3. This article provides a possible solution for the alignment among instructional approaches (authentic inquiry), technology design (AR) and learning experience.

Keywords: Augmented reality, Authentic inquiry, Practical skills

1. Introduction

Augmented reality offers a new form of interactivity between the physical and virtual worlds, and enhances users’ perceptions of the real world (Kesima & Ozarslan, 2012). According to Wu et al.’s review (2013), AR allows students to develop important practices, and has become one of the key emerging technologies in education. Although AR may present opportunities for teaching and learning, do students perceive an adequate level of realism when they are immersed in such a learning environment? How can researchers and educators work together to advance learning by aligning instructional approaches, technology design and learning experience? This study aims to propose a possibility to embed AR in authentic inquiry activities for contextualizing student exploration of medical surgery.

2. The AR learning modules

Authentic inquiry refers to performing the complex process which scientists actually carry out (Chinn & Malhotra, 2001). Instead of the simple inquiry tasks seen in most science textbooks, authentic inquiry tasks allow students to interact with computer-simulated experiments or equipment so as to develop their inquiry skills. AR is a promising way to combine authentic contexts and simulated experiments for student exploration. Therefore, we incorporated AR with authentic inquiry to engage students in two surgical procedures, laparoscopic surgery and cardiac catheterization. The authentic inquiry activities were designed to facilitate the students’ experience of the diagnosis of symptoms and to operate the laparoscopic surgery and cardiac catheterization simulators (Figures 1 and 2).

Originally, the simulated experiments in these two modules were designed for medical majors. For an outreach purpose, we invited senior high school teachers to design authentic inquiry activities for introducing medical practical skills to senior high school and helping them explore the possibility of their future career. A total of 32 senior high school students were grouped into eight groups, each of
which was provided with an Android tablet computer to interact with the simulators using the scanning function on the tablet. The tablet computer delivered an authentic context of a patient’s symptoms, and the students were required to diagnose the possible disease using the data such as X-ray images and electrocardiograms on the tablet. Then, the students used the simulator to help the patient recover from the disease. We applied the role-play technique in these activities whereby the students were told that they were surgery interns who were required to learn practical surgical skills. The students spent 2.5 hours working in a group to complete these two learning modules, laparoscopic surgery and cardiac catheterization.

A survey was conducted right after the modules to elicit the students’ perceptions of AR including the three constructs of realism, engagement, and motivation. Each construct consisted of six 5-point Likert-scale items modified from the study of Change, Lee, Wang and Chen (2010).

Figure 1. Laparoscopic surgery simulator. Figure 2. Cardiac catheterization simulator.

2.1 The laparoscopic surgery module

The students were told that they were to role-play surgery interns. Each group of students worked with a tablet and the simulator. First, a video clip on the tablet showed a patient describing his symptoms, and the students were required to diagnose the possible disease after checking the X-ray images. Second, some surgery options were shown to the students from which they were required to select the most appropriate procedure for the disease. Third, the students used the scanning function to activate the simulator and operated the laparoscopic surgery simulator which displays 3D images for promoting practical skills (Figure 1). Finally, the students needed to reflect on the pros and cons of laparoscopic surgery and to offer their opinions regarding improvements to such surgery.

2.2 The cardiac catheterization module

The cardiac catheterization module was developed following a similar procedure to that of the laparoscopic surgery module. The students role-played surgery interns to diagnose a patient’s disease from the electrocardiogram on the tablet. Then, the students used the scanning function to activate the simulator and carried out angioplasty by operating a catheter in the simulator. A 3D heart image was displayed synchronously on the computer to indicate the location of the catheter when the students moved it in the simulator. For some critical points, the computer provided a doctor’s advice to help the students overcome their difficulties operating the catheter. At the end, the students needed to answer some questions related to cardiac catheterization.

3. Findings and Conclusions

As Table 1 shows, the overall average score of student perceptions of AR was 4.1. The average scores of student perceptions of realism, engagement and motivation are separately 3.7, 4.3, and 4.3. We found evidence that embedding AR in authentic inquiry promotes students’ engagement and motivation in developing the practical skills for medical surgery. Although we used the same equipment as medical surgery, the students still perceived that the AR developed in this study was not as realistic as it might
have been. More improvements can be made to increase the realism of AR such as connecting surgical equipment to a human model which can react to students’ operation synchronously through sensors. Future studies need to investigate other important aspects such as examining the benefits of AR for learning practical skills in authentic inquiry activities.

Table 1: Summary of student perceptions of AR.

<table>
<thead>
<tr>
<th>Constructs</th>
<th>Mean</th>
<th>S.D.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Realism</td>
<td>3.7</td>
<td>1.1</td>
</tr>
<tr>
<td>Engagement</td>
<td>4.3</td>
<td>1</td>
</tr>
<tr>
<td>Motivation</td>
<td>4.3</td>
<td>0.7</td>
</tr>
<tr>
<td>Total</td>
<td>4.1</td>
<td>1</td>
</tr>
</tbody>
</table>

Acknowledgement

This research is partially supported by the “Aim for the Top University Project” of National Taiwan Normal University (NTNU), sponsored by the Ministry of Education, Taiwan, R.O.C. and the “International Research-Intensive Center of Excellence Program” of NTNU and Ministry of Science and Technology, Taiwan, R.O.C. under Grant no. NSC 103-2911-I-003-301.

References

Path Analyses of How Students Develop Conceptual Knowledge and Inquiry Skills in a Simulation-Based Inquiry Environment

Hsin-Yi CHANG*, Ying-Shao HSUb, Hsin-Kai WUb, & Chih-Ming CHENC

a Graduate Institute of Science Education and Environmental Education, National Kaohsiung Normal University, Taiwan
b Graduate Institute of Science Education, National Taiwan Normal University, Taiwan
c Graduate Institute of Library, Information and Archival Studies, National Cheng-Chi University, Taiwan
*hsinyichang@nknucc.nknu.edu.tw

Abstract: We implemented a simulation-based learning environment in a summer camp program in which 45 ninth-grade students conducted virtual experiments using computer-based simulations to learn the concepts of buoyancy. We collected data including all students’ responses to the pretests, assessments embedded during class in the learning environment, and posttests. All the tests measured the students’ conceptual knowledge of buoyancy and inquiry skills. We conducted path analyses to investigate how the students developed their conceptual knowledge and inquiry skills during and after learning within the learning environment. A common significant path is identified among students’ prior knowledge and skills and students’ developed knowledge and skills.

Keywords: Computer-based simulation, virtual experiment, inquiry, science, path analysis

1. Introduction

Computer-based simulations enable human-computer interactions that can allow students to conduct virtual experiments using simulations. Compared to hands-on experiments, virtual experiments are equal to or more effective for enhancing students’ understanding of scientific concepts due to their technological efficiency and their ability to provide ideal experiment conditions (Klahr, Triona and Williams, 2007; Zacharia, Olympiou and Papaevripidou, 2008). However, it is unclear how students develop their conceptual knowledge in a simulation-based virtual experiment environment as they conduct inquiries. Some research indicates that the development of conceptual knowledge and inquiry skills are often interwoven (Bao, Gotwals, Songer and Mislevy, 2006). Despite this, it is possible to measure conceptual knowledge and inquiry skills separately (e.g., P.-H. Wu, H.-K. Wu and Hsu, 2013), and to discern how conceptual knowledge and inquiry skills are interwoven at multiple time-points such as before, during, and after class. Specifically, we are interested in, in a simulation-based inquiry learning environment, how students’ prior knowledge and prior inquiry skills affect their knowledge and inquiry performance during class, and in turn after class in the posttests.

2. The Simulation-Based Inquiry Environment

The inquiry learning environment incorporates a newly developed simulation for students to conduct virtual experiments to learn factors related to an object sinking or floating in a given fluid. The students can change the values of four variables in the simulation to conduct their experiments (Figure 1): (1) the material of the object (duck): brick, wood, ice, styrofoam, or steel; (2) the size of the object (duck): large, medium or small; (3) the composition of the duck: solid or hollow; and (4) the type of fluid: water, saline water, gasoline, or mercury. In addition, on the upper right hand side of Figure 1, four virtual probes are provided so that the students can use them to measure the volume, mass, density or buoyant force of the object, and the density of the liquid. A small pop-up window appears on the left
side to indicate the volume of the fluid displaced after the duck is placed in the fluid. The students were asked to synthesize from their experiments, and to reason the variables directly related to the phenomenon of sinking and floating. In addition, we embedded a function in the simulation for the students to create a worksheet of their experiments recording the properties of the ducks and fluids they experimented with (Figure 2). Using the worksheets, the students were guided to reason that for a floating object, the buoyant force is equal to the weight of the object, and that for a sunken object (and a floating object), the buoyant force of the object in a fluid equals the density of the fluid times the volume of the object immersed in the fluid. Moreover, they were guided to draw their visualizations of the buoyant forces acting upon floating and sunken objects.

Figure 1. A screenshot of the bath duck simulation that allows students to conduct virtual experiments

Figure 2. A computer-generated worksheet recording the student’s experiments

3. Methods

The study involved 45 ninth-grade students who volunteered to participate in a summer science camp program at a public high school in North Taiwan. These students demonstrated high interest and motivation in learning science. They had not learned buoyancy prior to this study. The students spent 4 hours to complete their learning in the environment with the guidance of a science teacher. Each individual student took a pretest before and a posttest after. The pretest and posttest were identical and included two parts. The first contained 8 items to measure the students’ conceptual knowledge of buoyancy. The second part contained 15 items to measure the students’ inquiry skills including planning experiments, identifying variables, conducting reasoning, using evidence and evaluating explanations. The items went through several rounds of revision by science educators to ensure their content and construct validity. In addition to the pretest and posttest, another 8 conceptual items and 11 inquiry items were embedded in the learning environment to measure the students’ conceptual knowledge and inquiry performance demonstrated during the students’ learning in the environment.

We developed detailed scoring rubrics to score the students’ responses. In general, for the conceptual items, one point was given for an appropriate response and zero for an inappropriate one. For the inquiry items, two points were given for a high quality response, one point for a moderate quality response, and zero for a low quality one. Two independent raters coded all the tests, and the inter-coder agreement reached 95%. Inconsistent codes were discussed and resolved. Each individual student had six scores, namely, pretest knowledge, pretest inquiry, embedded assessment knowledge, embedded assessment inquiry, posttest knowledge, and posttest inquiry. We employed multiple regressions to conduct path analyses to test the relationships among these variables (Foster, Barkus and Yavorsky, 2006).
4. Results

We summarized the path analysis results in Figure 3 for the posttest knowledge scores and in Figure 4 for the posttest inquiry scores. Figure 3 shows a significant path from the students’ prior inquiry skills to their knowledge demonstrated during class, then to the inquiry skills demonstrated during class, and finally to their knowledge demonstrated in the posttests. The path shows direct significant effects from one variable to another among the four variables. The students’ prior knowledge did not have any significant effect. The students’ prior inquiry skills had an indirect effect on their inquiry skills during class, and the knowledge learned during class also had an indirect effect on the knowledge demonstrated after class. Figure 4 shows a similar path pattern for the inquiry skills demonstrated in the posttests. The only difference is that the students’ prior inquiry skills had a direct effect on their inquiry performance on the posttest.

Figure 3. A path diagram for conceptual knowledge demonstrated in the posttests

Table 1: Total significant effects for the path shown in Figure 3

<table>
<thead>
<tr>
<th></th>
<th>Prior Knowledge</th>
<th>Prior Inquiry</th>
<th>During Class Knowledge</th>
<th>During Class Inquiry</th>
</tr>
</thead>
<tbody>
<tr>
<td>Posttest Knowledge</td>
<td>0</td>
<td>0.045</td>
<td>0.087</td>
<td>0.307</td>
</tr>
</tbody>
</table>

Figure 4. A path diagram for inquiry skills demonstrated in the posttests

Table 2: Total significant effects for the path shown in Figure 4

<table>
<thead>
<tr>
<th></th>
<th>Prior Knowledge</th>
<th>Prior Inquiry</th>
<th>During Class Knowledge</th>
<th>During Class Inquiry</th>
</tr>
</thead>
<tbody>
<tr>
<td>Posttest Inquiry</td>
<td>0</td>
<td>0.816</td>
<td>0.112</td>
<td>0.392</td>
</tr>
</tbody>
</table>

We summarized the total effects on the posttest knowledge performance in Table 1, and on the posttest inquiry performance in Table 2. Table 1 indicates that the students’ inquiry during class had the strongest effect on their posttest knowledge performance, with their knowledge during class having an effect in between, and their prior inquiry skills having the least effect. For Table 2, the inquiry skills that the students possessed before class had the strongest effect on their inquiry performance on the posttest, with the middle effect of their inquiry developed during class, and least effect on their knowledge.
The red numbers in Figures 3 and 4 are the calculated error values for the dependent variables. These error values indicate variances not explained by the model. For both the knowledge and inquiry skills demonstrated during class, the error values are high, indicating that there are other variables not included in the model that might be better than the students’ prior knowledge and prior inquiry skills at accounting for their knowledge and inquiry demonstrated during class. We conjecture that the other variables include guidance and scaffolding from the teacher and learning environment. Qualitative analyses can verify this conjecture, but this is beyond the scope of this paper.

5. Concluding Remarks

Compared to SEM, multiple regression models for path analyses are suitable for studies with smaller sample sizes to explore possible relationships for further study. Through such techniques we found that compared to their prior knowledge, students’ inquiry skills are more important in terms of predicting posttest performance on both the conceptual and inquiry measures. Nevertheless, knowledge, especially the knowledge learned during class, still has effects on students conducting adequate inquiries. From a broader viewpoint, the development of knowledge and inquiry skills may be interwoven (Bao et al., 2006). When examining on a finer scale, the paths in the two diagrams follow a “Z” shape, starting from personal inquiry skills. Such a result has implications for the call for learning environments that emphasize the progressive development of inquiry skills. However, the results reported here are specific to the context of this study, including the simulation-based inquiry environment in which semi-structured curricular scaffolds and teacher guidance are provided to support student inquiry with the simulations (for details of the curriculum design, see Hsu, Chang, Fang and Wu, in press). Future studies can explore whether there are different paths when engaging students in other inquiry learning environments with less or more scaffolding or guidance.

Acknowledgements

This material is based upon work supported by the Ministry of Science and Technology, Taiwan, under grants MOST101-2511-S-003-055-MY3, and MOST102-2628-S-017-001-MY3. We would like to thank all the people who were involved in the collection and coding of the data.

References


Understanding middle and high school students' views of model evaluation and model change

Silvia Wen-Yu Lee*, Hsin-Kai Wu & Hsin-Yi Chang

*National Changhua University of Education
bNational Taiwan Normal University
cNational Kaohsiung Normal University
*silviawyl@cc.ncue.edu.tw

Abstract: This study aimed to understand students' views of the nature of model evaluation and the nature of change of models in different context. A total of 102 eighth graders and 87 eleventh graders were surveyed. Two cases, the SARS and dinosaur extinction, were presented to prompt students' ideas about different models proposed by scientists. The statistical results showed different context of the model influenced how the students viewed model evaluation and model change. The students’ answers also showed significantly differences between the high school level and the middle school level for their views of model change. The common reasons behind students’ choice were related to students’ understanding of the changeable nature of model and the science process. The students who chose that “one model is better than another” tended to justify their response by their understanding of the content. Interestingly, some students’ views of the dinosaur extinction model were guided by their beliefs that information about the dinosaurs is unfathomable. The findings suggest that researchers should be aware that the models chosen for teaching and for assessment can interact with other factors, such as their familiarity of the content, their level of education and understanding of the nature of science. The results from written responses were further used to develop a multiple-choice survey and validated in the follow-up study.

Keywords: views of modeling, computer-based survey, model evaluation, change of models

1. Introduction

Researchers found that students held little understanding of the concepts of models (Carey & Smith, 1993; Grosslight, Unger, Jay, & Smith, 1991; Saari & Viiri, 2003); even with formal training in modeling, students still encountered difficulties in fully understanding the nature of models (Harrison & Treagust, 2000; Schwarz & White, 2005). Researchers stated the needs of tapping into the interaction between epistemic beliefs and the contexts in which the epistemic beliefs being measured and being developed (Franco, Muis, Kendeou, Ranellucci, & Sampasivam, 2012). Earlier studies used interviews or paper-and-pencil questionnaires to understand students’ general beliefs of models and modeling. However, these studies provided students with little referential information to models. Thus the purpose of the study is to gain insight into the potential interaction of the views of models and the given context and the students’ justification to their views. This study focused on two of the major aspects of views of model, that is, the nature of model evaluation and the change of models.

In sum, we posted the following research questions: 1) What are the students’ view of model evaluation and model change in the two different context? 2) To which extent do the high school students’ views differ from the middle school students’ views in the given context? 3) How do the students justify their views of models in relation to the context?

2. Methodology
In this study, we surveyed 102 eighth graders and 87 eleventh graders. Two cases were presented to prompt students’ ideas about different models proposed by scientists. The first case involved two routes of infection for SARS (Severe Acute Respiratory Syndrome) virus, and the second case included different explanations of the dinosaur extinction. Students were asked to answer the multiple-choice questions and then provide a written response to justify their answers. We also asked students to rate that to which extent they were familiar with the two content of the two cases.

We conducted a series of Chi-square analyses including independent tests and goodness to fit tests for understanding the differences within the same educational level or between educational levels. We also used McNemar tests and McNemar-Bowker tests (Elliott & Woodward, 2006) for examining the consistency of students’ answers across different items. Opened coding methods were first applied to students’ written justification to their choice of answers. Then a list of coding schemes were tested on the data and modified until the coding schemes were saturated.

3. Findings

3.1 Model Evaluation

In terms of model evaluation, after reading the two cases, students had to make a choice among three options: (1) one model is better than another; (2) cannot know which model is better unless new evidence supports one of them; (3) both explanations can be valuable; there is no need to decide which model is better. Results show that nearly one fifth of the middle and high school students believed that one model is better than another. However, in the SARS case, nearly 70% of the middle school students and 57.47% of the high school students thought that both explanations can be valuable. The majority of high school and middle school students chose this answer for the SARS case (middle school $\chi^2(2) = 62.00, p < .001$; high school $\chi^2(2) = 23.24, p < .001$). For the dinosaur extinction question, the most chosen answer for high school students was “cannot know which model is better unless new evidence supports one of them (45.98%)” and “both explanations can be valuable; there is no need to decide which model is better (44.55%)” for middle school students. However, the results of chi-square analysis showed no statistical significant relationships between students’ educational levels and their views of model evaluation.

Further analysis with McNemar-Bowker tests also confirmed that the context of the item influenced students’ views of whether a model is better than another (p < .001 for middle school students; p = .004 for high school students). Only 50.4% of the middle school students chose the same answers between the two questions; even less percentage (40.2%) of the high school students had consistent answers between the two contexts. A high percentage of students who answered “both explanations can be valuable” for the SARS question shifted their views to “cannot know which model is better” when it came to the dinosaur extinction question. An interesting finding was revealed in students’ self-reported levels of understanding of the two topics. For middle school students, students reported similar level of understanding; however, for high school students, they reported significant higher level of understanding of the SARS topic than the dinosaur topic (SARS mean = 2.64; dinosaur mean = 2.38; p<.001). This could be a possible explanation of why the high school students seemed to shift their answers between the two contexts and believed that they could not know which of the models of dinosaur appeared to be better.

Overall, the high school students were more likely than middle school students to provide meaningful justification to their choices in both cases. Students who chose “one model is better than another” mainly focused their explanations on the science content of the cases (e.g., “if SARS were air-borne, then everyone should be infected by now”; “I think climate changes sound like the cause [for dinosaur extinction]”). Their justification to the answer of “no need to decide which model is better” focused on the changeable nature of models.
Examples of students’ responses included “models can change when the new one is better,” and “there could be more than one explanation or possibility.” Compared with students’ responses to the SARS case, the percentages of choosing the second option (cannot know which model is better unless new evidence supports one of the two) were higher in the dinosaur extinction case. Interestingly, one special set of responses to the dinosaur questions was unforeseen in the responses to the SARS case. Because “the dinosaur extinction happened long time ago; no one really knows” and “dinosaurs are dead”, many students believed that there is no way to know which model is better. We found that 43% of high school students who chose this option because that “dinosaur do not exist anymore” and only 18% of students who chose this answer really thought about the importance of finding new evidence. In the SARS case, students who chose the second option tended to justify their answers based on understanding of the scientific process (e.g., “if an error is found, scientists should correct it immediately”), science content (e.g., “I think it is air-borne”), or changeable nature of models (e.g., “if necessary, a model should change to respond to a new question”).

3.2 Change of Models
Based on the same SARS and dinosaur extinction context, we also asked students whether a model changes often. In the same context of the SARS and dinosaur extinction cases, students were asked whether models need to change often. Students could choose among the three options: (1) need to change often; (2) no need to change often; (3) it depends. For the SARS case, the most chosen answer was “it depends (44.55% for middle school; 55.17% for high school)”. A large percentage of students also chose “need to change often”. However, only 3.45% of the high school students chose “no need to change often” while nearly 15% of the middle school students preferred this option. There was significant relationship between students’ views of change of models and the two educational levels ($\chi^2 = 7.42, p = .024$) regarding the SARS question.

When answering the same question in the context of dinosaur extinction, nearly 52% of the high school students and about 41% of the middle school students believed that it depends. Less than 30% of the students chose either “need to change often” or “no need to change often” for both groups. The majority of students were unsure about whether models about dinosaur extinction need to change often. The results of McNemar tests showed that the context of the two questions had an impact on students’ views of whether models need to change, but only for high school students ($p = .002$). About 54% of the high school students held the same views of change of models between SARS and dinosaur questions. The percentage of the high school students who chose “no need to change” increased in the dinosaur case. For the SARS question, students who chose “it depends” or “need to change often” tended to provide reasons related the changeable nature of models (e.g., “there must be more than one pathway to spread the viruses” or “virus is always mutating”) and then scientific process (e.g., “it change when new evidence is found”). Students also gave similar explanations for the dinosaur question, but more students provided reasons related to the scientific process than changeable nature of model. A large percentage of the students to chose “no need to change often” did not provide a meaningful explanation. As for the students who provided justification, some of them stated “no need to change unless there is new evidence (coded as “science process”)” or “no need to the change the current model because multiple models can co-exist (coded as “ changeable nature of model”).

4. Discussion
Students’ views of model evaluation can be interpreted from a personal epistemological point of view (Justi & Gilbert, 2002; National Research Council, 2007). Based on different levels of personal epistemology (Yang & Tsai, 2010), the answer of “one model is better than another” is close to an absolutist perspective; the answer of “both explanations can be
valuable; there is no need to decide which model is better” is closer to a multiplist perspective; and the answer of “cannot know which model is better unless new evidence supports one of them” is similar to an evaluatist perspective. One interesting observation is that according to our data, students who chose an absolutist view of model seemed to focus on factual reasons. This can be interpreted as a way to support their judgment by evidence. Students who took a multiplist or evaluatist perspective, tended to think in terms of the nature of science (e.g., changeable nature of model).

In summary, we found that the different context of the model influenced how the students answered the questions of model evaluation and model change. The students’ answers also showed significantly differences between the high school level and the middle school level for their views of model change in the SARS case. The common reasons behind students’ choice were related to students’ understanding of the changeable nature of model or the science process. For the students who chose that one model is better than another tended to justify their response by their understanding of the content. Interestingly, some students’ responses to the dinosaur extinction case were guided by the beliefs that further information about the dinosaurs is unfathomable. These findings confirmed that students’ development of personal epistemology can be dynamic and somehow unstable influenced by factors such as the context, affection, or cognitive ability (Bendixen & Rule, 2004). The findings suggest that researchers should be aware that the models chosen for teaching and for assessment can interact with other factors, such as their familiarity of the content, their level of education and understanding of the nature of science.

5. References
Bio Detective: Student science learning, immersion experience, and problem-solving patterns

Mei-En HSU* & Meng-Tzu CHENG*
*Department of Biology, National Changhua University of Education, Taiwan
*mtcheng@cc.ncue.edu.tw

Abstract: Many studies have shown that there is a positive impact of serious educational games (SEGs) on student learning. Because of the game graphics, animations, sounds, and narratives, the learners can immerse in the virtual surroundings. Once they immersed, they might try their best to solve the in-game tasks. Therefore, the purpose of this study was to develop a SEG–Bio Detective and to evaluate its impact on student science learning outcomes. Moreover, we further investigated the relationships between students’ game immersion experience and their science learning and problem solving. The obtained results showed that student science learning can be significantly improved through Bio Detective play, but there were no significant correlations between game immersion and learning outcomes. Comparing the problem-solving patterns and problem-solving abilities between students with high- and low immersion experience, we found that students with high immersion experience had a more complete problem-solving pattern and a better problem-solving performance than students with low game immersion experience.

Keywords: serious educational games, problem solving, immersion

1. Introduction

With the rapid improvement of technology, video game play has become popular entertainment. Most of today’s children have the experience of playing video games. Due to the features of video games, such as excellent interaction and attractive entertainment, which engage players so much, researchers and educators commenced considering the probabilities of using video games in education since 2002 (Griffiths, 2002; Squire, 2008). Up to now, serious educational games (SEGs) have gradually become a term indicating any video games which are used for teaching and learning purposes in k-20 educational settings (Annetta, 2008). Mouaheb, Fahli, Moussetad and Eljamali (2012) suggested that playing SEGs is actually the process of learning and many studies have shown that SEGs did have a positive impact on various aspects of learning, such as learning achievement, cognitive development, learning motivation, and learning interests (Hwang, Yang, and Wang, 2013; Connolly, Boyle, MacArthur, Hainey, & Boyle, 2012). SEGs have been regarded as a potential vehicle which can enhances student learning because SEGs not only contain the entertaining features of games, but also combine those game features with learning content, which make learners engage in the game and further learn in the virtual surrounding (Van Eck, 2006; Cheng, Su, Huang, & Chen, 2013). It is generally argued that SEGs can provide learners with a vivid world which bridges virtual reality into reality in where players can experience so-called situated learning.

Brown and Cairns (2004) and Cheng, She, and Annetta (2014) contended that games can provide players with game immersion experience. The animations, sounds, and sophisticated graphics of games offer players an immersive environment in where they might ignore changes and forget everything at their surroundings; moreover, they might feel like they are the leading role of the game, and therefore put their whole concentration, thoughts, and even emotions into the game. While players experiencing game immersion, their intrinsic motivation and intrinsic rewards may increase, which subsequently make them actively play the game again and again. If the experience of immersion happened in the learning situation that allows learners to involve in the surrounding without external interference, then they will be willing to invest time and effort to learn. In terms of SEGs, the learners
will try to learn constantly because of the characteristics of SEGs of appropriately combining game features with learning content. Students will be getting familiar with the embedded concepts because of continuing practices while immersing in the game and better learning outcomes will subsequently reached (Liu, Cheng, & Huang, 2011).

On the other hand, the players have to learn the rules and plan and figure out ways to solve tasks and problems in the virtual environment in order to succeed in the game. Garris, Ahlers, & Driskell (2002) proposed the concept of input-process-outcome game model, claiming that SEGs offer an immersive environment which allows players to experience the game cycle of judgment–behavior–feedback and finally debriefing learning outcomes. The game cycle is actually the mechanism that provides the players with the complete problem-solving pathway. Therefore, it is argued that the problem-solving ability of players can be improved if they enjoy and immerse in the game (The Economist, 2005).

Although theoretical claims have been proposed by many researchers to explain why SEGs can improve student learning, there is still a lack of evidence that empirically investigates the impact of SEGs on student learning outcomes and how game immersion experience influences student learning outcomes and problem-solving patterns. Therefore, this study has a dual purpose. One is to develop an SEG, Bio Detective, to improve student science learning, and the other is to figure out the relationships between student science learning, game immersion, and problem solving through SEG playing.

2. Materials and methods

2.1 Serious educational game design

2.1.1 Bio Detective

In this study, the developed SEG, entitled Bio Detective, is an adventure/role-playing game. The storyline of Bio Detective is that there is a murder happened and the player should play as a detective’s assistant in the game. What he/she has to do is to collect clues, to conduct experiments to do blood-type and glucose tests, and to interpret data to find out whom the murderer actually is. Therefore, the scientific concepts embedded include the inheritance of blood-types (the inheritance of multiple alleles), blood-type test (antigen-antibody reaction) and glucose test.

Figure 1 is the laboratory scene in the Bio Detective. After collecting all the clues, the players are able to enter the lab to conduct experiments. They can drag any equipment they need to the table. If they drag the right equipment to the table, the button “Beginning of the experiment” will show up.

![Figure 1. Laboratory Scene.](image_url)

Figure 2 is the library scene. When the learners encounter problems and they don’t know how to solve them, they can go to the library to find information which might be useful in helping them solve the problems.
Bio Detective is created in an attempt to provide learners with a virtual environment as well as a vivid context in where they can actively learn the relative biology knowledge which is required for completing the game. It is hoped that students’ willingness and interests of learning can be improved because of their immersion in an atmosphere of suspense and excitement.

2.1.2 Learning objectives

According to the scientific concepts and science process skills embedded in the game, several learning objectives were addressed as below:

- **Cognitive domain** (the learner is able to)
  1. Comprehend principles and procedure of the glucose test
  2. Understand the principles and procedure of the blood-type test
  3. Interpret the experimental results and perceive the meaning
  4. Reinforce the scientific concepts of genes and blood types

- **Psychomotor domain**
  1. Promoting the learner’s ability to conduct the experiment independently
  2. Improving the learner’s ability to solve problems

- **Affective domain**
  1. Fostering the learner’s biological literacy
  2. Inspiring the learner’s enthusiasm and interest in exploring the field of biology

2.2 Participants

The participants were two classes of seventh grade students. One class consisted of 22 students and the other consisted of 30 students, resulting in a total of 52 students took part in the study.

2.3 Instrumentation

The instrumentation in this study includes the learning achievement assessment, the game immersion questionnaire (GIQ), a semi-structured interview guide and the students’ gaming performance.

In order to evaluate students’ learning outcomes, the embedded scientific concepts, the inheritance of blood-types and the principles and procedure of blood-type test and glucose test, are included in the learning achievement assessment. The assessment consists of fifteen multiple-choice items and five non-multiple-choice items (two are fill-in-the-blank questions and three are well-structured questions), and its total score is 100. The KR20 values of pretest and posttest were 0.75 and 0.78, respectively.

The GIQ was developed by Cheng, She, & Annetta (2014). It consists of 24 items categorized into three dimensions, engagement (9 items), engrossment (7 items) and total immersion (8 items), with
a five-point Likert scale ranging from 1(strongly disagree) to 5(strongly agree). The Cronbach’s alpha value of the total questionnaire and three dimensions were 0.94, 0.87, 0.87 and 0.95, respectively.

The semi-structured interview guide is designed to explore the students’ problem solving patterns while playing Bio Detective. It is divided into three parts. The first part includes questions regarding students’ perceptions of the in-game tasks (question 1 to 4). The second part consists of stimulated-recall questions, asking students what were they thinking and doing in the game using the stimulated recall method by viewing the recorded videos of their game play simultaneously (question 5 to 9). The last part is to investigate students’ feelings and suggestions regarding Bio Detective (question 10 to 13).

Finally, students’ gaming performance was recorded in the database, such as the number of times and duration students played the game and whether they have successfully completed the game or not. The gaming performance represents as the efficiency and accuracy of student problem-solving process.

2.4 Procedure

Before playing Bio Detective, students received a 45-min lecture about human ABO blood type first. After the lecture, the pre-test of learning achievement assessment was administrated.

Then, students were allowed to spend two sessions (90 min) playing Bio Detective. After finishing the play, an identical posttest of learning outcomes with different order of items was distributed to students no matter they succeed or fail the game. Moreover, the post-experience GIQ was also applied to investigate student game immersion after playing Bio Detective.

According to the GIQ scores, we selected four students with the lowest GIQ scores as low immersion group and the top four students as the high immersion group to conduct stimulated-recall interviews.

3. Result

3.1 Knowledge assessment

Table 1 shows the results of paired-sample t-test. It is demonstrated that there was no significant difference between student performance on pre- and posttest of multiple-choice questions (t=-0.15, p>0.05), but the performance on the non-multiple-choice test had significantly improved (t=-4.57, p<0.01). Overall, the students’ learning achievement were significantly improved after playing Bio Detective (t=-2.95, p<0.01).
Table 1: Result of the t test showing the per-test and post-test score.

<table>
<thead>
<tr>
<th></th>
<th>Pre-test</th>
<th></th>
<th>Post-test</th>
<th></th>
<th>t value (pre-post)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
<td>SD</td>
<td></td>
</tr>
<tr>
<td>Multiple-choice test</td>
<td>37.92</td>
<td>12.95</td>
<td>38.08</td>
<td>13.54</td>
<td>-0.15</td>
</tr>
<tr>
<td>Non-multiple-choice test</td>
<td>30.69</td>
<td>7.87</td>
<td>34.77</td>
<td>6.30</td>
<td>-4.57**</td>
</tr>
<tr>
<td>Total score</td>
<td>68.60</td>
<td>19.09</td>
<td>73.08</td>
<td>18.31</td>
<td>-2.95**</td>
</tr>
</tbody>
</table>

*p<0.01

3.2 Impact of immersion on learning achievement

Table 2 shows the results of Pearson correlations between game immersion experience and students’ learning achievement. It is showed that the three scales, engagement, engrossment and total immersion were highly inter-correlated. However, there was no significant correlation between game immersion experience and learning achievement.

Table 2: Result of Pearson’s Correlations between Game Immersion Experience and students’ learning achievement.

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>SD</th>
<th>Engagement</th>
<th>Engrossment</th>
<th>Total immersion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Engagement</td>
<td>32.06</td>
<td>7.36</td>
<td>0.46**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Engrossment</td>
<td>20.79</td>
<td>7.16</td>
<td>0.56**</td>
<td>0.71**</td>
<td></td>
</tr>
<tr>
<td>Total immersion</td>
<td>23.25</td>
<td>8.78</td>
<td></td>
<td></td>
<td>-0.15</td>
</tr>
<tr>
<td>Pre-test</td>
<td>68.60</td>
<td>19.09</td>
<td>0.23</td>
<td>0.00</td>
<td>-0.15</td>
</tr>
<tr>
<td>Post-test</td>
<td>73.08</td>
<td>18.31</td>
<td>0.25</td>
<td>0.02</td>
<td>-0.00</td>
</tr>
</tbody>
</table>

*p<0.01

3.3 Impact of immersion on problem solving pathway

Figure 4 reveals the high immersion group students’ problem solving pattern, inferring from the stimulated-recall interview. This shows that the high immersion group students knew what happened in the game. These students show that they had already planned how to find out the murderer before looking for the clues. They also demonstrated the ability to use the resources provided to search for information from the library or discuss with classmates when facing problems in the game. Most of them found the information provided in the library were useful. All of them completed the Bio Detective successfully.

![Figure 4. High immersion group students’ problem solving pattern.](image)

Note. KP= knowing the problem; PA=Plan in advance; CH=Collect hints; PD=Peer discussion; TAE=Trial and error; CD=Collect data; HP=Have problem; RP=Re-plan; GS=Games success.

Table 3 shows the results of the high immersion group students’ accuracy and efficiency of problem-solving ability. Student H1, H3, H4 completed the game in 40 to 45 min. Student H2 spend much more time than others. But his learning progress was better than any others in the high immersion group (pre-test score=46; post-test score=80).
Table 3: High immersion group students’ accuracy and efficiency of problem-solving ability.

<table>
<thead>
<tr>
<th>Student</th>
<th>Game times</th>
<th>Playtime</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>H1</td>
<td>2</td>
<td>41 Min.</td>
<td>success</td>
</tr>
<tr>
<td>H2</td>
<td>4</td>
<td>55 Min.</td>
<td>success</td>
</tr>
<tr>
<td>H3</td>
<td>2</td>
<td>45 Min.</td>
<td>success</td>
</tr>
<tr>
<td>H4</td>
<td>2</td>
<td>42 Min.</td>
<td>success</td>
</tr>
</tbody>
</table>

Figure 5 depicts the low immersion group students’ problem-solving pattern. It shows that the low immersion students might not know what happened in the SEG. These students required teacher’s guidance in order to understand the directions of the game. In the low immersion group, not all the students had a plan to find out the murderer. They tend to use trial-and-error strategy to solve problems. Some of them might ask teachers or peers for help when they cannot find the answer.

There was one student who couldn’t solve the problem, so she didn’t complete the game. While another student used guessing method to identify the suspects.

Table 4 shows the results of the students’ accuracy and efficiency of problem-solving ability low immersion group. Student L3 cost less time than any other. However, she used trial and error method to complete the game. Like other students in the low immersion group, she also didn’t have any plan before looking for clues, instead, she find those clues one by one.

Student L1, L3, L4 spends 55 to 65 min to finish the SEG. There was one student who failed to complete the game.

Table 4: Low immersion group students’ accuracy and efficiency of problem-solving ability.

<table>
<thead>
<tr>
<th>Student</th>
<th>Game times</th>
<th>Playtime</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>L1</td>
<td>2</td>
<td>59 Min.</td>
<td>success</td>
</tr>
<tr>
<td>L2</td>
<td>4</td>
<td>65 Min.</td>
<td>failure</td>
</tr>
<tr>
<td>L3</td>
<td>1</td>
<td>28 Min.</td>
<td>success</td>
</tr>
<tr>
<td>L4</td>
<td>3</td>
<td>55 Min.</td>
<td>success</td>
</tr>
</tbody>
</table>

Overall, the high immersion group students demonstrated higher level of problem-solving ability than students in the low immersion group. In the high immersion group, all of them completed the SEG and spend less time compare to the low immersion group.

4. Discussion and conclusions

In recent years, many studies have shown that SEGs can provide students with meaningful learning experience, as the design of SEGs attempts to combine learning content with game format which increases the opportunity of motivating and engaging students in the learning activities embedded in the game (Federation of American Scientists, 2006). In other words, proper game design promotes
immersive experience of students and increases their willingness to learn the concepts and materials in the game (Cheng, She, & Annetta, 2014). Today, SEGs have been considered one of the potential methods for students to learn and construct knowledge (Pivec, 2007). In this study, a SEGs, Bio Detective was developed and its effectiveness was investigated. It is found that the participants did learn the scientific concepts embedded in Bio Detective as their performance on the knowledge assessment significantly improved after playing the game. The results are in alignment with previous studies illustrating that learning through playing SEGs can be effective for student biology learning (Cheng, Annetta, Folta, & Holmes, 2011; Cheng, &Annetta, 2012; Cheng, Su, Huang, & Chen, 2013; Cheng, She, &Annetta, 2014).

What is interesting in the study is that, students performed significantly better on non-multiple-choice questions rather than multiple-choice ones. As we might know that multiple-choice questions generally involve tasks related to recognition, and students all have to do is to choose or recognize the knowledge that has been learned. On the contrary, non-multiple-choice questions require students to recall relative information from their memory. Compared to recognition, it is a higher-level information processing because students have to recall specific knowledge and concepts without any clues or hints (Sternberg, 2009). Only when the knowledge has been elaborately processed by students, can they been easily and accurately recalled. As a result, it is more difficult for students to answer non-multiple-choice questions if they are not very familiar with and never elaborately process the concepts. However, our study revealed that the use of Bio Detective is much more helpful for students to understand what they have learned and even learned better because of their better performance on the recall tasks.

This study also explored whether the different degree of student game immersion can impact student science learning outcomes through SEG play. The results of Pearson’s Correlations indicate that the three dimensions, engagement, engrossment and total immersion were highly inter-correlated. However, there was no significant correlation between game immersion experience and science learning outcomes. The results are pretty much in accordance with the previous study conducted by Cheng, She, & Annetta (2014). Researchers have suggested that cognitive load might be a key that should be considered while learning through SEG play (Cheng et al., 2014; Cheng et al., 2013). What should be taken into account is, in which aspect do the students really invest their mental effort? The game itself such as storylines and narratives or the educational aspect such as the embedded concepts and knowledge required for completing the game? Obviously, more mental efforts students spend in the game, more engaged they are. However, they might ignore the educational materials they should learn in the SEGs because of over immersing in the game.

Furthermore, immersion experience engages and even absorbs learners in a situation. While experiencing game immersion, learners invest much time and efforts in solving the tasks because of the enhancement of their internal motivation. Hence, the ability of problem solving should be considered an important element of learning outcomes of SEG play either. Only assessing the effectiveness of SEGs from the aspect of concept acquisition might underestimate the impact of using SEGs on students’ science learning. The obtained results of the study additionally demonstrate that although there were no correlations between game immersion and concept learning, differences in student problem solving patterns did exist between students with high and low game immersion experience. Sternberg (2009) proposed that people who have better problem solving ability will prefer to spend time on planning how to solve the problem. On the contrary, people with poor problem solving ability will cost lots of time to trial and error, without any strategy. According to the playtime, game accuracy and problem solving pattern, we found that students with high immersion performed better than the students with low immersion in the game. The high immersion group would plan how to complete the task before finding the clues and they spent less time completing Bio Detective than students in the low immersion group did. Namely, in the high immersion group, students have better problem solving performance. When the students immersed in the SEG, the immersive experience might enhance learning. Students would change their perspective and engage in the game (Dede, 2009) as if they were the real detective. That is why high immersion group had better problem solving performance.

Problem solving is an important aspect of learning. Students learn at school and the teachers provide lots of knowledge for students to learn. It is hoped that students will utilize what they have learned in the classroom to solve daily life problems instead of rote learning. As a pilot study, we found that students with high immersion experience perform better problem-solving strategies in the SEG –
Bio Detective. Although SEGs provide learning experience, the problem solving ability which is affected by SEGs need to be further investigated.

Acknowledgements

The authors would like to thank the National Science Council, Taiwan, R.O.C. for financially supporting this research under Contract No. NSC 101-2511-S-018 -004 -MY3 and NSC 102-2815-C-018-018-S.

References


The effect of students’ effectiveness and attitude in heterogeneous and free grouping cooperative learning applied in sixth-grade students’ Scratch program teaching

Lei Chen*, Xiuyu Yang*, Xintong Wang*, Feng-kuang Chiangab*
*School of Educational Technology, Faculty of Education, Beijing Normal University, China.
Collaborative & Innovative Center for Educational Technology, Beijing Normal University, China.
*fkchiang@bnu.edu.cn

Abstract: Although the method of cooperative learning has been used popularly, but researchers paid less attention to different grouping way in the past, especially heterogeneous and free grouping. In this paper, researchers analyzed students’ effectiveness and attitude in heterogeneous and free grouping cooperative learning in sixth-grade students’ Scratch program teaching. 52 students participated in the study and the Scratch animation works assessment scale was prepared and revised by some experts, students and Scratch teaching teachers. Many interesting things have found in the study, such as heterogeneous grouping brings better students’ works, and more efficiently arouses students’ curiosity. In addition, free grouping stimulates more positive attitudes, while heterogeneous grouping leads to more learning pressure, and boys have less pressure in Scratch learning than girls. This can provide some constructive suggestions for cooperative learning.

Keywords: heterogeneous grouping, free grouping, Scratch, program teaching

1. Introduction

Nowadays, with the rapid development of modern information technology, technology plays a critical role in the education reform, and the ability of ICT has been regarded as essential skills for primary and middle school students. Since Ministry of Education published the curriculum guidelines of primary and middle school in 2003, education of information technology has been a major issue. And at once setting forth from the third level, students deliver a single lesson of information technology every week at least, thus as to improve the ability of ICT. Though students in primary can learn many things from lessons of information technology, programming lessons are abandoned or delayed until middle or senior high school.

Scratch is a new programming environment developed by MIT Media Lab. It is suitable for children aged over eight years old to make their own small programs creatively, such as interactive stories, animation, games and art (Chen, 2009). He (2013) also said that students could make a program easily by dragging blocks, which is suitable for young children. Some teachers have used the method of cooperative learning, but they don’t give enough attention to it, and whether different grouping has some close relationship with students’ effectiveness is still unknown.
2. Literature review

2.1 Study of Scratch programmed instruction methods

Xie (2013) puts forward a teaching method “creating situation—analyzing cases—interacting between teachers and students—design (personalized works) —sharing and communicating” by exploring the teaching model and the method of Scratch. Xu & Huang (2013) found that collaborative project-based on learning could enhance students’ effectiveness and attitudes toward Scratch. With the didactic and inquiry teaching methods respectively, Yang (2010) tried teaching the Scratch program design, and the results showed that with these two kinds of teaching methods on the fifth grade primary school students, there were not significantly different, and the teaching method and the mathematics learning achievement didn’t have close relationship. Ke (2013) explored the potentials of computer-aided and mathematics game making activity based on Scratch on promoting students’ mathematics learning. Research showed, participants who involved in making computational game were more positive to mathematics learning. And the experience and design of game could connect to daily mathematical experience.

In summary, students usually make a Scratch work in a group, whose aim is to express their own feelings by engaging in developing comprehensive abilities, cultivating creative thinking, and learning how to cooperate with others (Brennan & Karen, et al, 2010). So the method of cooperative learning is suitable for Scratch learning. However, there are few studies about Scratch program cooperative learning of pupils. So researchers tend to focus on this area.

2.2 Studies of free grouping and heterogeneous grouping

Zhang (2006) made a study based on heterogeneous grouping, and after a semester’s study the experimental group whose members were divided into groups by their basic ability got higher scores than the control group. The research showed that the heterogeneous grouping could improve different aspects of ability. A research (Deng & Huo, 2010) also found the basketball skills of heterogeneous grouping had been improved much more than that of free grouping. Teaching activities may be carried out in accordance with students’ aptitude.

All in all, there are no definite conclusions about whether different grouping in programmed learning will bring disparate imparts on students’ effectiveness and attitude. Similarly heterogeneous and free grouping are often used during daily teaching activities, but studies about students’ effectiveness based on heterogeneous and free grouping are few.

3. Method

3.1 Participants

52 students in two classes in the 6th grade of Elementary School in Beijing are included in the research. The classes were assigned as the experimental group (22 students, Class 1) and a control group (30 students, Class 3). For 8 weeks, the same Scratch program teaching was conducted in both two groups. Only one researcher taught them Scratch.
3.2 Research design

Despite the lack of pre-test, two classes are normal classes, which aren’t selected and changed. Given that students had no significant differences in effectiveness and attitude at first, one class (class 1) is free grouped, and the other one (class 3) is heterogeneous grouped. One group has two students. Then through evaluation of the works and an attitude questionnaire survey of cooperative learning groups, researchers compare the Scratch programming learning effectiveness and attitudes of students grouped by two different ways.

According to the objectives of the study and relevant literature information, the framework of this research was established as follows.

![Figure 1. Framework of Research](image)

3.3 Course design

According to the researchers’ teaching practice and the general process of cooperative learning in the programming previously submitted by Ju (2007), the teaching model during the research is shown in the following figure:

![Figure 2. Programming cooperation model (Source: revised from chart of Ju, 2007)](image)

During the eight-week research with one lesson a week each class, designed by the author, the teaching content was divided into four sections: broadcast, asking-answering, logical operators, and making Scratch works. Instructor, also a researcher, taught Scratch in two pilot classes in the same way.

As shown in the figure, cooperative learning in each class has the following main procedures: Based on the pupils’ previous knowledge, the teacher designs, reasonable and practicable teaching and learning activities, presents problems of cooperative learning and tasks after solving the knowledge problems. In task, with learned knowledge, students design Scratch programming theme, analyze problems and determine the division after group discussion, and solve problems by group
cooperation. Then the teacher and students, show group works, evaluate performance of another group and present improvements together. At last students' summary what they have learnt in this class.

3.4 Research tools

3.4.1 Scratch animation works assessment scale and learning attitudes questionnaire of Scratch

The assessment scale was developed by the researcher and revised by two experts, two masters and five Scratch teaching teachers across the country, consisting of four dimensions: design, aesthetic, technology and innovation. The scores of Cronbach’s Alpha are both more than 0.8.

The questionnaire consists of five dimensions, respectively, motivation, self-efficacy, usability, usefulness, autonomous learning and learning pressure. It was developed by researchers referring to Professor Huang Guozhen's questionnaire of learning attitude and scientific accept mode (Hwang, Tsai & Tseng, 2010). Score of Cronbach’s Alpha in one class is 0.895, and another is 0.933.

4. Results

4.1 Analysis of students’ animation works

At the end of the term, works of each group were collected, but some group did not hand in their works finally. There were twenty-two animation works in total, nine from class 1 and thirteen from class 3. According to Scratch animation works assessment scale, three Scratch researchers (undergraduate students), gave a score for each animation seriously. The average score of three undergraduate students was the final score of animation works. Because each class had less than 30 students, and some students did not hand in their works, so samples might be not enough. When analyzing the samples, the method of nonparametric analysis was used to analyze works between class 1 and class 3 comparatively.

4.1.1. Heterogeneous grouping might bring better students’ works

By nonparametric analysis of animation works between class 1 and class 3, it indicated that these two kinds of grouping has no significant differences in scores of works. But According to descriptive analysis, it indicated that the average score of students’ works in class 3 was about 79.0 points and the average score of students’ works in class 1 was about 83.7, four point seven points higher than works in class 3.

Table 1: Descriptive Data of students ‘works in different dimensions

<table>
<thead>
<tr>
<th>category</th>
<th>class</th>
<th>N</th>
<th>Mean</th>
<th>Std. Deviation</th>
<th>Std. Error Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>total</td>
<td>1</td>
<td>9</td>
<td>83.72</td>
<td>6.45</td>
<td>2.15</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>13</td>
<td>78.97</td>
<td>6.14</td>
<td>1.70</td>
</tr>
<tr>
<td>design</td>
<td>1</td>
<td>9</td>
<td>21.20</td>
<td>1.47</td>
<td>0.49</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>13</td>
<td>19.66</td>
<td>2.11</td>
<td>0.59</td>
</tr>
<tr>
<td>aesthetics</td>
<td>1</td>
<td>9</td>
<td>20.39</td>
<td>2.00</td>
<td>0.67</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>13</td>
<td>20.25</td>
<td>1.57</td>
<td>0.43</td>
</tr>
</tbody>
</table>
Table 2: The results of Nonparametric analysis of students’ works in different classes in different dimensions

<table>
<thead>
<tr>
<th>Null Hypothesis</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>The distribution of total is the same across categories of class</td>
<td>.03*</td>
</tr>
<tr>
<td>The distribution of design is the same across categories of class</td>
<td>.014</td>
</tr>
<tr>
<td>The distribution of aesthetics is the same across categories of class</td>
<td>.65</td>
</tr>
<tr>
<td>The distribution of technology is the same across categories of class</td>
<td>.03*</td>
</tr>
<tr>
<td>The distribution of innovation is the same across categories of class</td>
<td>.13</td>
</tr>
</tbody>
</table>

Then researchers analyzed the scores in four dimensions of Scratch animation works assessment scale, found that only the dimension of technology had significant differences between the two classes. And the average score of technology in class 1 was about 21.5 points, 1.9 points higher than that in class 3.

By nonparametric analysis on items of the dimension of technology, it indicated that score of the item “Scratch can arouse my curiosity” had significant differences between the two classes. And The score in class 1 was 0.5 points higher than that in class 3.

4.2 Analysis of students’ attitudes towards Scratch

4.2.1 Free grouping brought out more positive attitudes.

All participants completed the attitude questionnaire towards Scratch at the end of term, and forty effective questionnaires were collected finally, 27 from class 3 and 13 from class 1. By nonparametric analysis of attitudes between class1 and class 3, it indicated that these two kinds of grouping had significant differences about attitudes towards Scratch. Descriptive statistics showed that the score of attitudes in control class(class 3) was 16.5 points higher than that in experimental class (class1), and the score in each dimension of attitudes in class 3 was higher than class 1. So students in free grouping had more positive attitudes towards learning Scratch than students in heterogeneous grouping.

Table 3: Descriptive Data of students’ attitudes in different class in different dimensions

<table>
<thead>
<tr>
<th>category</th>
<th>class</th>
<th>N</th>
<th>Mean</th>
<th>Std. Deviation</th>
<th>Std. Error Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>total</td>
<td>1</td>
<td>13</td>
<td>40.77</td>
<td>19.76</td>
<td>5.48</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>27</td>
<td>57.26</td>
<td>20.73</td>
<td>3.99</td>
</tr>
<tr>
<td>motivation</td>
<td>1</td>
<td>13</td>
<td>6.77</td>
<td>3.52</td>
<td>0.98</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>27</td>
<td>9.30</td>
<td>5.10</td>
<td>0.98</td>
</tr>
</tbody>
</table>
4.2.2 Heterogeneous grouping might lead to more learning pressure.

The scores in the six dimensions of learning attitude questionnaire of Scratch were analyzed, and the result shows that the dimensions of usability, usefulness and pressure had significant differences between the two classes, especially learning pressure between two classes had abnormally significant differences (sig=. 00**).

Table 4: The results of Nonparametric analysis of students’ works in different classes in different dimensions

<table>
<thead>
<tr>
<th>Null Hypothesis</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>The distribution of total is the same across categories of class</td>
<td>.01*</td>
</tr>
<tr>
<td>The distribution of motivation is the same across categories of class</td>
<td>.011</td>
</tr>
<tr>
<td>The distribution of self-efficiency is the same across categories of class</td>
<td>.014</td>
</tr>
<tr>
<td>The distribution of usability is the same across categories of class</td>
<td>.04*</td>
</tr>
<tr>
<td>The distribution of usefulness is the same across categories of class</td>
<td>.03*</td>
</tr>
<tr>
<td>The distribution of autonomous is the same across categories of class</td>
<td>.11</td>
</tr>
<tr>
<td>The distribution of pressure is the same across categories of class</td>
<td>.00**</td>
</tr>
</tbody>
</table>

Then researchers analyzed the five items of learning pressure detailedly and found that three of them had significant differences between the two classes. These items included “Usually I can’t concentrate on learning Scratch.”, “I'm stressed when using Scratch.”, and “It takes me much time to master the use of Scratch.”

A descriptive analysis was carried out, and reverse scoring was used in the dimension of learning pressure when analyzing data in SPSS, so it means the more score of learning pressure is, the less pressure students have. From Table 3, it indicated that the score in each item of pressure in class 3 was higher than that in class 1, so students in class 3 have less learning pressure than that in class 1. In other words, students in heterogeneous grouping feel more stressed than students in free grouping. Table 3 shows the score in each item of learning pressure in class 3 was respectively higher than that in class 1. So heterogeneous grouping makes students feel more stressed. They couldn’t concentrate on learning Scratch, felt stressed when using Scratch and needed to spend more time to master it.
4.3 Analysis of gender impact on student attitudes

The scores in six dimensions of learning attitude questionnaire of Scratch were analyzed and it shows there were no significant differences totally between boys and girls in two classes. However, boys and girls had abnormally significant differences in pressure, and boys had less pressure than girls in Scratch learning. The results of nonparametric analysis of students’ attitudes between boys and girls in different dimensions are shown in table 5.

Table 5: The results of nonparametric analysis of students’ attitudes between boys and girls in different dimensions

<table>
<thead>
<tr>
<th>Null Hypothesis</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>The distribution of total is the same across categories of class</td>
<td>.09</td>
</tr>
<tr>
<td>The distribution of motivation is the same across categories of gender</td>
<td>.78</td>
</tr>
<tr>
<td>The distribution of self-efficiency is the same across categories of gender</td>
<td>.91</td>
</tr>
<tr>
<td>The distribution of usability is the same across categories of gender</td>
<td>.20</td>
</tr>
<tr>
<td>The distribution of usefulness is the same across categories of gender</td>
<td>.19</td>
</tr>
<tr>
<td>The distribution of autonomous is the same across categories of gender</td>
<td>.17</td>
</tr>
<tr>
<td>The distribution of pressure is the same across categories of gender</td>
<td>.00**</td>
</tr>
</tbody>
</table>

5. Conclusions and discussion

5.1 Analysis of students’ works

Students in heterogeneous grouping got more scores than students in free grouping, which means heterogeneous grouping may lead to more positive effect in effectiveness than free grouping. Also, students in class 1 had significant differences in technology, especially in the item of “Scratch can arouse my curiosity”. So Scratch teaching in heterogeneous grouping could more efficiently arouse students’ curiosity and stimulate a strong desire to learn.

5.2 Analysis of students’ attitudes

According to the nonparametric analysis of attitudes between class 1 and class 3, it indicated that different grouping made significant differences about attitudes towards Scratch. Descriptive statistics show that the score of attitudes in free grouping was 16.5 points higher than that in heterogeneous grouping, and the score in each dimension of attitudes in free grouping was higher than heterogeneous grouping. Free grouping brings out more positive attitudes.

Dimensions of usability, usefulness and pressure had significant differences between the two classes, especially learning pressure (sig=.00). From Table 3, it indicated that students in heterogeneous grouping feel more pressure than students in free grouping. In other words, Heterogeneous grouping may lead to more learning pressure, particularly in these aspects: concentrating on learning Scratch, using Scratch and costing more time to master it.
Boys and girls have abnormally significant differences in the dimension of pressure both in two classes, and boys have less pressure than girls in Scratch learning. Maybe girls have difficulty in programming, and some studies have confirmed this inference.

5.3 Research Limitations

This study analyzed students’ effectiveness and attitude in heterogeneous and free grouping cooperative learning in sixth-grade students’ Scratch program teaching, and some distinguished results have been shown above. However, without pre-test, it may have some influence on the later analysis of the results. The research design will be improved and more differences in heterogeneous and free grouping are going to be found in the future.

References


Xiaotong Xu, & Yaping Huang. (2013). Study on Scratch collaborative learning achievement of different students.


Yinting He. (2013). Analysis on the effects of Scratch program teaching based on guided learning.

Designing Mobile Application for STEM: Building Individual Interest and Supporting Creative and Innovative Thinking Skills

Ilker YENGIN

Institute of High Performance Computing, A*STAR, Singapore
yengini@ihpc.a-star.edu.sg

Abstract: This paper explains the design of a mobile application that is aimed to support the development of creative and innovative skills using stories as the external factors to create and maintain students’ long-term individual interests in STEM area. In this paper, first, we discussed the connection between individual interests and creative – innovative thinking process. Then, we described research proved pedagogical methods, which based on using stories to build and maintain students’ individual interests. Following that we illustrated the design of mobile application that applies the pedagogical methods into practice. In the discussion part, we discussed the potential use of the application as well as future follow-up research studies.

Keywords: STEM, creativity, creative thinking, innovative thinking, mobile learning, interest building, individual interest, mobile application design.

1. Introduction

Creative and innovative thinking are couple of the key thinking skills that would be core to science, technology, engineering and mathematics (STEM) education (Atkinson and Mayo, 2010). Developing students’ creative thinking skills have significant role in building modern economies. Investments on creativity and innovation had a dramatic positive impact on the global competitiveness of countries (MacLeod, et.al., 2007). Similarly, research on highly creative and innovative countries (e.g. Finland, Norway etc.) showed that supporting creative and innovative thinking clearly facilitates the transition from the postindustrial economy, towards an emergent knowledge-based creative economy (Duell, Wright and Roxburgh, 2014). To remain competitive in a challenging global environment, supporting creative and innovative thinking in work and learning plays a crucial role (Ferrari, Cachia and Punie, 2009).

Specially designed work and learning spaces encourage us to think in creative and innovative ways (Oksanen and Ståhle, 2013). According to a meta-analysis of 32 research studies on creativity, enabling social interaction is identified as one of the significant category of the key characteristics of conditions promoting creativity and innovation (Davies et.al, 2013). Similarly, it is possible to provide same characteristics into STEM education to support creativity and innovation (Cooper and Heaverlo, 2013). For instance, using technology to provide social understanding may have different positive influences on creative and innovative thinking skills in STEM (Kärkkäinen and V.Lancrin, 2013). However, solely providing technological tools are not sufficient to develop creative and innovative thinking skills by itself (Selwyn, 2011; Jackson et.al., 2012). It is suggested that technology should often be perceived as a catalyst for change in the application of pedagogical methods in order to have successful results (Watson, 2011).

One of the pedagogical methods that help to have successful results is to helping students to develop an individual interest (Arnone et.al., 2011). For instance, increasing individual interests on a particular subject/object enhance creativity (Shalley, Zhou and Oldham, 2004; Grant and Berry, 2011). Developing students’ individual interests towards the STEM related learning activities help them to increase their chance to develop creative and innovative thinking skills (Bairaktarova and Evangelou, 2012).
It has been suggested that students with individual interest for STEM subjects are likely to be motivated to pursue the STEM careers that require creative and innovative thinking (Tyler-Wood, Knezek and Christensen, 2010). Developing interests in STEM requires carefully planned interactive activities in school environment. Informing students on the topic is not a sole component of developing interest (Hidi, Renninger and Krapp, 2004). There should be activities supporting both positive feelings and opportunities for gathering knowledge for social understanding in order to be able to shift students’ interests on STEM areas (Renninger and Shumar, 2002). Instead of simply supporting having fun and participation in science, institutions should also provide activities for productive social understanding by addressing individual interests (Stocklmayer and Gilbert, 2002). Research of U.S. National Research Council (2006) clearly indicated that students shows more productive participations in STEM activities when they developed individual interests by finding meaning thorough social understanding in the learning strategies. Social understanding develops through the process of observation, introspection, and imagination (Carpendale and Lewis, 2004).

Student’s individual interests in STEM may not be influenced all the time by the school related activities (Hidi, 1990). First of all identifying students’ individual interests would need a great effort. Second, these interests may not find a place to be pointed in school subjects. Thus, it is important to understand what types of activity may address interests of different range of the student population. Interest on STEM subject might be an outcome of interactive activity or a content in which the activity and the environment may have an effect on the dynamics of the interests (Krapp, 2005). Other external motivators such as role models (e.g. scientists, teachers) and peers may contribute to an increased emerging individual interest (Krapp and Lewalter, 2001).

After individual interest is developed well, students usually have a tendency to pursue further understanding independently and work on developing deep understanding in STEM fields (Renninger, 2000). It is expected that external factors allowing the well-developed individual interest make it possible for students to maintain positive feelings and interests for STEM related activities in future. Thus it is very important to create a learning environment that provides set of social activities within the inclusion of knowledge from the external individuals in order to create interests in STEM. Also engaging social activities with an established individual interest may extend creative and innovative thinking as suggested in several studies (Stocklmayer and Gilbert; Tyler-Wood, Knezek and Christensen, 2010).

As mentioned above paragraphs, it is expected that individual interest support creative and innovative thinking. Also external motivators could increase it. Thus building individual interests for students becomes a very important factor in success of students in STEM. Addressing the needs, this study discusses the design of a mobile learning application using a pedagogical model that implements core dynamics of individual interest building using the external motivators in order to develop creative and innovative thinking in STEM field.

The pedagogical methods that are using external motivators to increase individual interest in STEM are discussed in the first part of the paper. Second part of the paper illustrates the design of the mobile learning application, which is conceptualized according to the pedagogical methods in the first part. Finally a discussion part explains the possible use case scenarios of designed application in STEM education.

This paper would be useful for the instructional designers, teachers and school administrators who like to learn more about developing creative and innovative thinking skills and individual interests in STEM field as well as an example design of mobile application for this purpose.

2. Pedagogical Methods

2.1 Increasing Individual Interest

Students with high interest in STEM subjects may possibly have gained more knowledge than their peers due to the fact that they may put more effort on the STEM activities (Tobias, 1992). This implies that students with prior knowledge on the STEM area may have increase individual interest. The argument is also supported by the research. Research suggested that as students become more familiar with an area, their interest toward the areas is expected to increase (Tobias, 1994).
The external sources of information could be integrated in STEM education to increase the prior knowledge of the students. Especially, external institutions (e.g. science centers, research companies, universities and museum etc.) and individuals (e.g. scientists, innovators, artists and researchers etc.) may be a great external resources of information. As the research shows, connecting and having a regular relationship with these institutions and individuals in an out of school may develop a personal interests slowly over time with a tendency of having a long term effect on students’ knowledge and values on the subject (Ainley and Ainley 2011; Dabney et. al., 2012; Renninger, Hidi and Krapp 2014). Thus, following the research based suggestions; it is assumed that when students start connecting to institutions and individuals in and out of school, they would have greater prior knowledge on the STEM related areas, which would also increase individual long-lasting interests on STEM area.

Based on this, it is possible to have institutions and individuals as external factors to increase the individual interests. As a pedagogical method, we could seek for possibilities to establish connection with external institutions and individuals. In the application design, we implement elements to apply this practice to increase students’ individual interests.

2.2 Using Stories

Stories create engaging and pedagogically effective learning experiences (Mcquiggan et.al., 2008; Kuyvenhoven, 2005; Dyer and Wilkins, 1991). Stories are very effective in learning and theory building (Badreddine and Buty, 2011). Using stories increase students’ interests on a particular subject (Fulmer and Frijters, 2011). Similarly, stories are used as pedagogic practices in social learning to support science education (Shelby and Ernst, 2011). Stories that are interesting motivate students to continuously to keep their attention on the subject. For instance, students reading stories with elements such as novelty, character identification, life themes, and activity level may have a continuous interest to get more knowledge about the topic (Hidi, 2001; Hidi and Baird, 1986) including science (Sandoval, 1995). Also stories are powerful applications with a potential to creating an emotional connection to the individual interests that results extending with some degree of continuity on interaction (Tan, 2013).

Also stories enable students to make connections between ideas in the ongoing STEM issues which help them to have a personal meaning and interests towards the science concepts (Scott, Mortimer and Ametller, 2011). Stories are also allowing teachers and students to have a platform to share the experiences of the individuals in STEM area to help students to grasp the related concepts, accommodate with diverse perspectives and realities of the experts in the field (McDrury and Alterio, 2003).

It is an accepted idea that enabling students to write stories are commonly used to support their development of the creativity (Hennessey and Amabile, 1988). Similar to writing stories, it is suggested that reading stories are also supporting students to have some initial ideas to be able to have some creative associations (Smogorzewska, 2014). Thus, the system design is expected to have an impact on students’ creativity with a certain degree. Reading stories approach is expected to grab and maintain the students’ attention. Especially some interesting elements such as life themes of scientist and how they react to the challenges in daily life, their solutions and services to the important problems of global scene etc. is expected to arouse the students’ curiosity which will also expected to turn into the individual interests with a long term interactions. Thus, carefully selected stories of external individuals in STEM area may help students to understand the tasks involved in the professional STEM related fields such as decision making process in daily tasks, theories, tools and work environments. As the second pedagogical method, we implement the idea and key elements of using stories to create individual interest in STEM related areas.

3. System Design

Implementing the pedagogical methods described in Section 2, we designed a learning environment where students have a chance to use a platform to connect to the science and read stories about the problems in science and the scientists’ lives. As research studies suggested, by connecting students to science using the stories as a meeting point, it is expected to maintain students long-term individual interests on STEM areas. Also as an effect of continuous long-term individual interest, it is expected
that students may develop their creative and innovative thinking skills, which would be essential in STEM education. To enable the pedagogical methods, we designed the system architecture prototype as in Figure 1. Accordingly, story content collector engine collects the STEM related stories form different online resources using RSS then sort and tags into database. Distributor engine sends out the stories to student based on the rules defined by individual interest profiler. Individual interest profiler creates a dynamic profile of individual interest for each student in the system and feeds the distributor engine with rules. Interactive interface displays the content received from the distributor engine with a set of interactive events.

![Figure 1. Mobile Application System Architecture](image)

3.1 Story Collection and Distribution

Story contents are fed into the system in forms of newsletters and short articles forms, which are collected from the online resources. The content collections are categorized in a variety of topics. Themes would be challenges in science, new scientific and technical developments as well as professional lives of scientists.

The content collection is centralized and processed by the story content collector engine using “Really Simple Syndication” (RSS). The content collector engine signs up automatically to the various online resources related to STEM subjects and collects the content to centralize them before the distribution. Once the system administrator set up the resources’ addresses for the content collection the content collector will continue to function for as long as the system runs. System administrators have no need to check the online resources to find out what is going on and whether there is a new update available. All the updates and stories come automatically to the content collector engine and sorted and tagged according to the topics.

The distributor engine sends out the stories to students according to a rule set defined by the individual interest profiler. When students log in to the system, they also start building their profile in the system. According to their profile, students would regularly receive stories related to STEM fields, which are selected and personalized according to each student’s individual interests in the topic. The selection of the stories according to individual interest analysis is described in the next section.

3.2 Individual Interest Profiler

The individual interest profiler identifies students’ individual interests. The individual interests and related mechanism is dynamically calculated according to students’ previous selections of the topic as
well as asking feedback from the students. When students registered into the system they are asked to fill a registration form. In the registration form, students are asked to respond to a questionnaire developed according to the instruments that measure the students’ views on nature of science (Osborne, Simon and Collins, 2003; Chen, 2006). Furthermore, students are also asked to indicate which topics and subjects they feel they like to read in STEM areas. Moreover, in addition to the initial data collection, after each time reading a story, students are asked to give feedback on their individual interest in the stories. Students indicate their self-interest using a 5 points rating scale. Our system does not rely on self-rating data, it also have its own tracking and measurement system embedded in individual interest profiler. Accordingly, the system tracks and measures students’ interactions by activity analysis, clicks and patterns of behavior in relation to other interface and content.

3.3 Interactive Interface

Interactive interface displays the story content. This interface is where students engage in to the stories. It is designed for mobile devices including tablets, which allow students to benefit from the application in and out of the school and to follow the content according to their own phases. The interactive interface is designed as a scientist’s diary application (e.g. design of look and feel of the application, activities etc.) where students would get their STEM stories regularly. In the application students have a library where he could select stories that are updated according to his/her individual interests regularly (Figure 2).

![Figure 2. Mobile Application Interactive Interface for reading STEM related stories](image)

The interface has basically functions as e-readers application. However, to have a look specific to STEM education, the reading interface provides additional design functions to create experiences for the students to help them feel like they are engaging in scientific diary reading experiences. This special design makes the interface design different than the traditional e-readers. More than just embedding the content from the online source and delivering it on a tablet device without adding any value to the content, the interface have interactive features such as scrolling up and down in a blocks of text and using some artwork to resemble a science related reading activity. In the application, the reading interaction is scrolling based in which students need to scroll up and down to see the contents and swipe it right and left to turn the pages. In that sense, the interface captures the essence of reading science diary rather than reading a book. Adding the interactivity, the interface creates a meaningful science reading experience for the students. Interface also has standard functions such as search, bookmark, sharing, printing annotation and taking notes.
4. Discussion and Future Studies

The mobile application intended to build students’ self interest, which has a direct connection in supporting the creative and innovative thinking process. Using the mobile application to deliver science related stories as the external resources for motivational process; it is aimed to increase the individual interest in STEM areas. Relying on the research studies that show the connection between the interest and creative – innovative thinking, we expect that the mobile application will be helpful to support the development of that mentioned skills of creative and innovative thinking. The design of mobile learning application was based on the implementations of pedagogical methods that we picked based on the suggestions of several reach studies as explained with details in the previous chapters. This mobile application can be used to increase students’ long-term interaction in STEM subjects. As the application is good for tablet devices, students may use it in and out of the classroom. As a future study, we expect to test the mobile application in terms of design effects on helping students in developing motivation and the creative – innovative skills within experimental studies.

5. Conclusion

In this study, using the stories in STEM areas as the external motivators, we designed a mobile application that helps building and maintaining students’ individual interest in STEM area. As previous studies suggested it is expected increasing interest would directly support students’ creative and innovative thinking. For addressing this issue, applying the pedagogical models using stories as the external motivators, we provided a design for a mobile application. The mobile application is expected to build and maintain students’ individual interests, which also result in development of their creative and innovative skills in STEM area. The pedagogical methods and the design of the mobile learning application are discussed in the course of the paper. Finally in the discussion part, the possible use case scenarios and future studies are mentioned shortly.

This paper would be useful for the instructional designers, teachers and school administrators who like to learn more about developing creative and innovative thinking skills and individual interests as well as an example design of mobile applications for this purpose.

References

A systematic literature review.


Shelby, A., & Ernst, K. (2013). Story and science: How providers and parents can utilize storytelling to combat anti-vaccine misinformation. *Human vaccines & immunotherapeutics, 9*(8), 1795-1801.


The Development and Evaluation of an Educational Game- Shimmer© with Computer Visualization for Optics Learning

Huei-Tse HOUa*, Shu-Ming WANGb, Hsin-Hung YUA & Shie-He LIUA

aGraduate Institute of Applied Science and Technology, National Taiwan University of Science and Technology, Taiwan
bDepartment of Information Management, Chinese Culture University, Taiwan

*hthou@mail.ntust.edu.tw

Abstract: Learning physics involves the reasoning of abstract, dynamic, and complex that cannot be mapped to learners’ daily experience or observed with naked eyes. with the advancement of computer technologies, nowadays, computer technologies, such as simulation and scientific visualization, are able to depict the scientific phenomena with vivid visualization and rich interaction. Recently, in addition to provide simulation and scientific visualization, many educational games were developed to support learning of various subjects in order to promote the level of students’ engagement in learning. Educational games with clear goals, immediate feedback, and adequate scaffolding were considered as a powerful means of improving students’ learning motivation and subsequent learning performance. Though game-based learning has been a trend in recent years, researchers suggested that research on the application of game-based learning to STEM is still relatively limited. To fill the literature gap, this study developed a game – Shimmer© to support the learning of optical concepts. This study conducted an experiment with 50 seventh-grade students. An analysis of pretest and posttest was conducted to examine the effects of the game on learning. In addition, the participants’ evaluation of game and their gaming experience were assessed via using flow constructs. Furthermore, individual differences were addressed as well. Results showed that the participants generally better capture the concepts of light reflection and refraction in visual way. Besides, further analyses of individual differences indicated that boys and girls as well as students with high and low prior knowledge evaluated the game differently. Implications for the results of this study are to be used as guidance for subsequent game development and design of instructional strategies.

Keywords: Game-based learning, STEM, Optics, flow

1. Introduction

Scientific learning involves numerous abstract concepts, which cannot be directly observed or mapped to learners’ day-to-day experience (Anderson & Barnett, 2013). For example, the concepts of optics and wave cannot be directly observed without adequate instruments (Hamam, 2003). When learning these concepts, it would be difficult for learners to capture and integrate the abstract, complex, and discrete ideas to form correct conception of a particular phenomenon. Sometimes, learners would even form misconceptions if improperly guided (Anderson & Barnett, 2013; Masson, Bub, & Lalonde, 2011). Nonetheless, with the advancement of computer technologies, nowadays, these abstract, dynamic, and complex phenomena can be depicted by using computer simulation or scientific visualization techniques to improve learners’ understanding (Chen, Pan, Sung, & Chang, 2013; Van Dam, Forsberg, Laidlaw, LaViola, & Simpson, 2000). Previous studies have adopted technologies, e.g., virtual reality or interactive computer simulation, to support the teaching of abstract concepts, such as electrostatics (Anderson & Barnett, 2013), object motion (Masson et al., 2011), and optics (Mzoughi, Herring, Foley, Morris, & Gilbert, 2007).
As to learning optics, it could be difficult for learners to catch the concepts of optics, as it is invisible to naked eyes (Mzoughi et al., 2007; Van Dam et al., 2000). Though the lab experiment could be helpful to improve students’ understanding, it could be complex and costly for setting up a lab experiment as the class is getting larger. This situation makes computer visualization be an ideal tool to support the learning of the concepts of optics (Mzoughi et al., 2007; Van Dam et al., 2000). In response to this issue, Mzoughi et al. (2007) developed a 3D interactive computer graphics system – WebTOP for teaching and learning optics. WebTOP adopted the features of visual simulation to illustrate the optical concepts, such as reflection and refraction, lasers and scattering etc., allowing that learners better understand these phenomena. The system can support both classroom use and self-guided study and received generally positive affirmation from both educators and students. However, one general constraint for computer simulation is that users might be lack of motivation to engage in the repetitive process or they could easily get bored. This situation could diminish the educational benefits of using computer simulation for teaching and learning.

To improve users’ motivation and level of engagement in learning, employing an educational game to support teaching and learning has been a trend in educational practices nowadays (Anderson & Barnett, 2013). An educational game adopts game mechanism in addition to computer simulation and visualization. Like computer simulation, computer games allow players to interact with the content and receive immediate feedback, which can be used to adjust their action or conception accordingly. Well-designed computer games provide clear goals with different levels of challenge that require players to explore the means to achieve. Rewards are provided as players reached the goals afterward. This process provokes intriguing experience that promotes players’ motivation to and engagement in gaming (Mayo, 2009; Prensky, 2001; Young et al., 2012). In the educational context, games can be adapted to individual learners’ pace and scaffold them through the learning process. For example, complex learning tasks can be decomposed to smaller tasks at the beginning of a game. As the level of challenge gradually gets easier, players would have opportunity to practice simple tasks several times before they deal with more complex learning tasks (Mayo, 2009). Through the repetitive practice, learners are expected to attain the mastery of particular cognitive skills (Prensky, 2001).

Witnessing the surging popularity of games and advantages of adopting games for learning, many educational games have been developed in recent years. However, there is relatively limited research on game-based STEM (science, technology, engineering, and mathematics) in the past decade (Young et al., 2012). In order to fill the literature gap, this study developed an educational game – Shimmer© to support the learning of optics and conducted a preliminary study to improve our knowledge about adopting game to support learning. The description of Shimmer© is presented in Research Method section.

To explore why and how people play games, previous game-related studies have used the idea of flow from psychology literature to assess the optimal gaming experience (e.g. (Barzilai & Blau, 2014; Hsu & Lu, 2004; Kiili, 2006). Flow experience refers to the perception of being totally absorbed to the activity in which people engaged. People would feel a sense of enjoyment, distortion of time, loss of self-conscious, and intrinsically motivated when they were in flow state (Csikszentmihalyi, 1994). Flow is therefore adopted as an important indicator of intriguing experience when playing games. Nonetheless, there are prerequisite conditions for the occurrence of flow experience. One essential element of flow is that the level of challenge has to match the player’s level of skills. Otherwise, the player could get anxious when the level of challenge is significantly higher; on the contrary, the player would get bored when the level of skills is significantly higher. Moreover, people need to feel in control of what they are doing and feel that they are pursuing a clear goal (Kiili, 2006). This study adopted flow experience to assess the participants’ evaluation of Shimmer© and their gaming experience. Besides, the effects of individual differences have been regarded as influential factors in educational research (Li, Cheng, & Liu, 2013; Yukselturk & Bulut, 2009). This study further looked into the plausible effects of individual differences on participants’ learning outcomes, evaluation of game, and gaming experiences. Concluding from the above, the purposes of this preliminary study are summarized as follows:

1. To examine the difference of learning outcomes before and after playing the game - Shimmer©.
2. To explore the plausible differences of game evaluation and gaming experience of Shimmer© for students with different gender.
To explore the plausible differences of game evaluation and gaming experience of Shimmer© for students with different level of prior knowledge.

The initial findings of this preliminary study could help us better understand the effectiveness of Shimmer©, which visualizes the abstract and dynamic concepts of optics. These findings would also serve as a guideline for improving the future version of Shimmer©. Moreover, based upon the findings, suggestions for incorporating scientific visualization with game-based learning were purposed. The following sections are to delineate the introduction of game and the experimental design.

2. Research method

2.1 Procedure and Participants

This educational game used in this study was a 3D simulation game - Shimmer©, which was developed via Unity3D game engine. In the game, the player needs to protect civilians in a small village from the invaders. In each stage, there were a laser cannon, several plane mirrors and lenses, and one invader sat on a floating island as shown in figure 1. By clicking on the mirrors and lenses, players could calibrate their angle and location. Once the players finished the calibration, they could click on the laser cannon to launch a laser and check how it traveled. The goal of the game was to correctly set up the angle and location of the lenses so that the laser could shoot the invader. At current version of the game, players were allowed for unlimited trials. The concepts of optics introduced in the current version of Shimmer© are about reflection and refraction. At the beginning stages, there were plane mirrors only. Players could simply calibrate the angle of plane mirrors to achieve the goal. These stages were designed to help players to be acquainted with the game and learn the basic ideas of light reflection. At the further stages, convex and concave lenses as well as obstacles sited on the path of laser transmission were added, which would gradually increase the difficulty level of the game (as shown in figure 2). There were 15 levels of game challenges in current version of Shimmer©. A non-player character (NPC) - the Sage was created to provide relevant knowledge of light refraction. Players could click on the Sage to read the needed information at any time during the game. Screenshot of the provided information is shown in figure 3.

Figure 1: Game screenshot of the beginning stage.
The participants of this study were 50 seventh-grade students in a junior high school in northern Taiwan, including 28 males and 22 females. The experiment was conducted in a PC classroom. Before the experiment, students were asked to fill a pretest to assess their prior knowledge of light fraction. Students have to complete the pretest in seven minutes. The pretest was followed by a five-minute introduction of the game. Afterward, students were to freely play the game for 20 minutes. Finally, students were asked to fill a posttest as well as a flow instrument. The total time of the experiment was around 45 minutes.

2.2 Instruments

This study developed a test, which was used in the pretest and posttest session. The test included 8 multiple-choice questions, which were primarily adapted from the textbook, and 4 questions that asked students to draw the direction of light transmitted through the lenses. The example of drawing question is as shown as figure 4. In this example, the arrow showed the direction of the light. Participants were asked to draw the travel path of the light. Students were given one point for each correct answer.

Figure 4: Example of a drawing question

The flow instrument was adapted from Kiili (2006). 22 items were used to assess two major components of the flow construct. The first component is flow antecedents, which represents players’ evaluation of the game in terms of the level of challenge, goal, feedback, control, and playability. The second component is flow experience, which reflects players’ perception of concentration, time distortion, autotelic experience, and loss of self-consciousness when playing game. The flow instrument was measured using five-point Likert scale from 1 (strongly disagree) to 5 (strongly agree).

3. Data Analysis and Results

3.1 Learning outcomes

SPSS 20.0 was used to analyze the collected data. Table 1 summarized the participants’ scores in pretest and posttest as well as the results of mean difference test. As shown in Table 1, overall, there was a significant difference between the pretest and posttest. However, further analysis revealed that there was no difference in the scores of multiple-choice. Nonetheless, after playing the game, students’ performance of drawing question was significantly higher than before (t = -5.14, p < 0.001).

Table 1: Summary of students’ performance in pretest and posttest.

<table>
<thead>
<tr>
<th>Type of questions</th>
<th>Pretest (n=50)</th>
<th>Posttest (n=50)</th>
<th>t-statistics</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>S.D.</td>
<td>Mean</td>
<td>S.D.</td>
</tr>
<tr>
<td>Multiple-choice</td>
<td>4.38</td>
<td>1.92</td>
<td>4.46</td>
<td>2.11</td>
</tr>
<tr>
<td>Drawing</td>
<td>0.46</td>
<td>0.68</td>
<td>1.18</td>
<td>1.02</td>
</tr>
<tr>
<td>Overall</td>
<td>4.84</td>
<td>2.24</td>
<td>5.64</td>
<td>2.58</td>
</tr>
</tbody>
</table>

For the flow construct, the Cronbach’s α of flow instrument was 0.971, suggesting high reliability (Nunnally, 1978). The participants generally possess positive evaluation of the game. Specifically speaking, the participants evaluated Shimmer© as challenging (mean = 3.64), with clear goals (mean = 4.10), providing prompt feedback (mean = 3.71), controllable (mean = 3.64), and playable (mean = 3.52). Meanwhile, all the means of sub-constructs of flow experience exceeded 3.5, suggesting that the participants generally have positive flow experience while playing game.

3.2 Individual differences

Individual difference was considered as an important factor in educational research (Price, 2006; Yukselturk & Bulut, 2009). In the context of gaming, the stereotypical view of males is that they
might be more interested and engaged in playing video game than females (Chumbley & Griffiths, 2006; Yee, 2006). In addition, when playing game, males and females might behave differently. For example, girls would be more focus on causal interaction with other players, whereas boys would tend to focus on game-related conversation (Young et al., 2012). These differences would affect their evaluation of an educational game. In this regard, this study conducted a further analysis to explore the potential differences between boys and girls in evaluation of Shimmer© in terms of flow antecedents. The results showed that boys, in contrast to girls, showed higher evaluation of game challenge (t = 2.96, p < 0.05), feedback (t = 3.62, p < 0.01), and playability (t = 2.72, p < 0.05). However, there were no differences observed in their evaluation of game goal (t = 1.69, p > 0.05) and control (t = 1.855, p > 0.05). Regarding the flow experiences, boys reported higher degree of flow experience in concentration (t = 2.22, p < 0.05), autotelic experience (t = 2.78, p < 0.05), and loss of self-consciousness (t = 2.51, p < 0.05). Nonetheless, the difference in time distortion was not observed (t = 1.71, p > 0.05).

To further explore the plausible differences in students’ flow experience, this study conduct a test for difference between the means of students with high and low prior knowledge. Based on their scores in the pretest of this study, students whose scores were above the top 27% were categorized into high prior knowledge group (N=14), whereas those who below the bottom 27% were categorized into low prior knowledge group (N=14). The results indicted students with high prior knowledge reported higher degree of experiencing time distortion than those with low prior knowledge (t = 2.28, p < 0.05). However, there were no differences observed in other flow dimensions. The interpretation and discussion of these results are to be presented in the following section.

4. Conclusion and subsequent research

With the technology advancement, modern computer games are able to simulate the scientific phenomena and create an immersing environment that follows the natural laws of physics for educational purposes (Anderson & Barnett, 2013). With this affordance, computer games have been used to support learning of various subjects. In particular, computer games are commonly used to support the learning of abstract concepts or complex procedures (Kebritchi & Hirumi, 2008). This study developed an educational game – Shimmer© to support the learning of optical concepts. The initial findings suggested that Shimmer© was helpful to improve students’ conceptions of reflection and refraction, particularly in visual depiction. Participants generally reported positive flow experience, which was considered as an essential component of intriguing gaming experience, when playing Shimmer©. The discussion and implications of findings are delineated as below.

First of all, the game – Shimmer© helped the participants better capture the ideas of reflection and refraction. In particular, after playing the game, the participants were more able to correctly draw the path that light travels among mirrors, convex and concave lenses than before. Shimmer© allowed players to form their hypotheses of how light travels and then set up the angle and location of lenses to freely test and alter their hypotheses for unlimited times. This process could enhance students’ autonomy in learning, which could promote their learning motivation and thus achieve better learning outcomes (Ryan, Rigby, & Przybylski, 2006; Vogel et al., 2006). Nonetheless, there was no significant difference in multiple-choice questions between the pretest and posttest. One plausible explanation may result from the form of problem representation in the game. In Shimmer©, students need to calibrate the lenses in order to make the light hit the target. In the process, they were able to observe the results of light transmitted through lenses of different angles and locations. Therefore, they would form the conceptions of reflection and refraction in visual way. Shimmer© didn’t require players to memorize factual knowledge to achieve the game goal. In other words, students needn’t to answer the traditional form of test when playing game. Thus, their performance on multiple-choice questions could be limited. Similar notions have been proposed by previous research. Masson et al. (2011) suggested that the engagement benefits of games are suited for training particular cognitive skills. Nonetheless, this immersing environment and repetitive process might make learners be less accomplished with regard to traditional test performance (Young et al., 2012).
Secondly, analysis of gender difference showed that boys tend to have higher evaluation in most dimensions of flow antecedents. As boys were regarded as typical gamers, they might be more able to adapt to game control than girls are (Chumbley & Griffiths, 2006; Yee, 2006). However, lacking of control could impede players from immersing in the game play, which in turn diminishes the level of engagement and the effects of game-based learning (Hou, Wang, & Tsai, 2013; Scoresby & Shelton, 2011). Nonetheless, in this study, there were no differences in boys and girls’ perceptions of game goal and controllability. This may result from the game goal of *Shimmer©*, whose goal was to shoot the invader with a laser cannon, which was simple and clear enough for both boy and girl participants. Moreover, the participants only need to click on the lenses to calibrate the angle and location of them. No complicated operations were required. Regarding the flow experiences, boys seemed being more absorbed into the game as they generally reported higher degree of flow experience in concentration, autotelic experience, and loss of self-consciousness. However, the difference in sense of time distortion between boys and girls was not observed. This non-significant result could be attributed to the limited time of the game session (20 minutes) and a small sample size. Subsequent research is suggested to extend the game session and conduct a larger scale of experimentation to further investigate the gender differences in flow experience.

Lastly, regarding to individual differences, though the participants with high and low prior knowledge generally reported positive flow experience (i.e. mean of all sub-dimensions exceeded 3.5), results indicated that participants with higher prior knowledge tend to have significantly higher sense of time distortion. It seemed that they were more engaged in playing game. This finding was probably because students with higher prior knowledge were able to capture the ideas in game. Thus, they would be more focused on solving the problem than those with lower prior knowledge were. This finding also emphasized the importance of the match between the level of challenge in game and the skills of players, which is a prerequisite condition for the occurrence of flow experience (Csikszentmihalyi, 1994). Thus, it is suggested that the design of an educational game should consider the prior knowledge of learners in order to induce the optimal gaming experience while playing (Barzilai & Blau, 2014; Kiili, 2006).

5. Research Limitations and subsequent study

As a preliminary study, this study was limited in the time of game session and sample size; thus, these initial findings should be interpreted with cautions. First, the short game session could be a limitation for the students to properly evaluate the game and to have flow experience. In this manner, the longer game sessions are needed to further investigate the effects when using the game – *Shimmer©* to support the learning optical concepts. Secondly, an elaborated experimental design with more participants, which involves the control group provided for other instructional practice such as simulation, would help us better compare the effectiveness of game-based learning with that of traditional practice. Lastly, the advantages of game-based learning are not merely to promote learners’ engagement. Mayo (2009) suggested using in-game assessment to track the sequences of players’ behaviors, such as the number of attempts to solve a problem, the timing of seeking online help, or interaction among players. This approach can help instructors to further look into the learning process and provide feedback for game developers to improve the game design. In this manner, the subsequent research is suggested to employ sequential analysis to explore players’ gaming patterns, which would help us better understand how players interact with the game and improve the future version of *Shimmer©*.

Acknowledgements

This research was supported by the projects from the National Science Council, Republic of China, under contract number MOST-102-2511-S-011-001-MY3, MOST-100-2628-S-011-001-MY4, MOST-103-2511-S-034-001 and NSC-102-2511-S-034-001.
References


A Three-Stage Augmented-Reality-Facilitated Earth Science Instructional Process for Dispersing Learning Style Differences

Chang-Hwa WANG*, Cheng-Ping CHENb*
Department of Graphic Arts and Communications, National Taiwan Normal University, Taiwan
Department of Visual Communication Design, Taipei College of Maritime Technology, Taiwan
*chenjp0820@yahoo.com.tw

Abstract: Studies have proven that merging hands-on and online learning can result in an enhanced learning experience. However, the effects of learning styles have substantially affected online learning performance. As ICT continues to develop, an augmented reality (AR)-embedded instructional orientation could provide additional hands-on experiences to the classroom. In contrast to traditional online learning, multiple in-classroom activities may be involved in an AR-embedded e-learning process, thus could reduce the effects of individual differences. Using a three-stage AR-embedded comprehensive instructional process, an experiment was conducted to investigate the influences of student’s learning styles. The results of the study showed that overall learning achievement was significant for the AR-embedded instruction. Nevertheless, as no significant difference found among different learning styles, indicating that our multiple activities oriented AR learning process may have helped disperse the effect of different learning styles.

Keywords: augmented-reality-facilitated learning, learning style, science learning, interactive learning environment

1. Introduction

Some researchers asserted that students learn more effectively with e-learning environments because students like interactive learning that provided by recent interactive technologies (Lee, Choi, & Park, 2009; Hatziapostolou, & I Paraskakis, 2010; Ali Karime, Hossain, A. S. M. M. Rahman, Gueaieb, Alja’am, & El Saddik, 2012). Hrastinski (2009) indicated if learner has an opportunity to control their learning environment, they would have more interest and willing to learn in classes. In an interactive e-learning environment, students would become more positive and active.

Augmented reality (AR) is one of such interactive technologies. It mixes virtual and real world by means of displaying virtual objects onto real images in accordance with target triggers (markers) that manipulating by users. In addition to visualization, users can interact with virtual objects (Chehimi, Coulton, & Edwards, 2007). Many studies revealed that AR systems have educational values because students enjoyed the interaction with virtual objects which is also effective to improve students’ learning performance. Among various interactive technologies, interactions with AR were found to be particular helpful for learning spatial concepts (Kaufmann & Schmalstieg, 2003; Kirner, Zorzel & Kirner, 2006; Juan, Beatrice, & Cano, 2008). Researches have also indicated that learning style is important in laying the groundwork for understanding students’ learning performance, especially for e-learning, in which learner characteristics is necessarily adapted to the interactive instruction (Huang, Lin, & Huang, 2012). Related works have been done with learning styles in relation to learner’s participation, learning quality, and performance of e-learning (Shaw, 2012; Marković & Jovanović, 2011). The results of these researches exhibited there are significant relationship between learning styles and e-learning outcomes in general. Nevertheless, none of these discussions went beyond the realm of traditional online e-learning.

The adaptation of individual differences to e-learning has been discussed for several decades. Johansen and Tennyson asserted that adaptive advisement could help students in perceive knowledge in learner-controlled, computer-based instruction (Johansen & Tennyson, 1993). Magnisalis, Demetriadis,
and Karakostas (2011) claimed that Artificial Intelligence and Web 2.0 techniques could support collaborative learning in an online learning environment. However, all these techniques involve complicated computer algorithms and likely need to comply learning activities within a computer monitor, therefore lack of direct human interactions. Unlike traditional online learning, multiple activities may be involved in an AR-facilitate e-learning process. An AR-facilitate e-learning is able to take place in regular classroom settings, instructional activities may include life lectures, student manipulations of virtual objects, peer discussions, and even written exercises after experiencing the AR. These multiple activities could be more adaptive for learners with different learning styles. However, little research has been done on this issue.

The present study assumed that a comprehensive AR-facilitate learning process consists of various types of learning activities, therefore there will be less effects of learning styles on learners' achievements. A quasi-experiment was performed to examine the effects of learning styles on learning achievement while the comprehensive AR-facilitate learning process was given. In the present study, a comprehensive AR-facilitate learning process includes lecture, hands-on AR experience, peer discussions, and written exercises.

2. Related Work

2.1 AR-facilitate Instruction

Azuma (2009) first recognized the AR as a technique that link between the real and virtual world. Yuen, Yaoyuneyong, and Johnson (2011) gave an up-to-date definition to AR. In their definition, AR has three distinctive characteristics: (a) it is the combination of real world and virtual elements, (b) it is interactive in real-time, and (c) it is registered in three dimensions. Thus, AR has some potential to influence instruction and learn knowledge from different fields.

Several researches have used AR systems in education, including mathematics, science, language, and medicine. For example, in their experiment, Kirner, Zorzal and Kirner (2006), a “Game of Word” used plates containing symbols of English alphabets, when setting up a word completed by the plates in front of the webcam, the related virtual object appears over it. They believed that this game was able to motivate the users to interact and create solutions in an attractive AR environment. In addition, Juan, Beatrice and Cano (2008) presented an AR system for learning the interior of the human body. Learners were able to “open” the abdomen of a virtual human body using their own hands. Learners also saw inside the human body virtually, and observed the areas where the stomach and the intestine are located. More recently, Matsutomo, Miyachi, Noguchi, and Yamashita (2012) created a real-time visualization system, which can visualize a composite image of source materials and their generated magnetic field utilizing the AR technique. They claimed that with such a system, electromagnetics learners can observe the magnetic distribution in a virtual real-time manner.

More AR-facilitate science learning researches have been done in this decade. Recent discussions of instructional applications of AR have gone beyond the effectiveness of the AR per se. For example, in order to better understand the effective strategies that are appropriate for AR-facilitate learning Yoon, Elinich, Wang, Steinmeier, and Tucker (2012) compared four conditions for learning science in a science museum using AR and knowledge-building scaffolds. Results indicated that students demonstrated greater cognitive gains when scaffolds were used. The limitation of above research findings is that they only viewed AR learning as a standalone activity and fall short to vision the entire AR-facilitate learning process, including lecture, peer discussions, and other classroom activities, as a whole. Wang and Chi (2012) demonstrated a comprehensive AR-facilitate learning process, emphasizing in-classroom interactions, to teach fundamental earth science for junior highs. The entire learning process included teacher’s lecture, hands-on AR experiences, peer discussions, and written exercises. They thought that AR-facilitated instruction could improve the understanding of spatial concepts and be easier to acquire the course contents. Nevertheless, the differences between individual students were not discussed in this study. They suggested that further research on individual differences, for example, the learning styles is necessary.

A Synthetic work of AR research was done by Bujak, Radu, Catrambone, MacIntyre, Zheng, and Golubski (2013). They reviewed recent research on AR learning. They highlighted the potential benefits and limitations of using AR to deliver learning experiences, by presenting an analysis based on
psychological constructs, and by comparing AR applications to physical and virtual manipulatives. They concluded that although AR shows great promise for extending the resources used for educating our students, there is much research to be done. Finally they suggested that researchers must more specifically address the usefulness of AR from a psychological perspective.

2.2 Learning Styles

Viewing from the psychological perspective, a line of research has found that learner characteristics had great effects on learning performance. Lamia and Mohamed Tayeb (2013) recognized that learning styles, thinking styles, and levels of knowledge and abilities are key learner characteristic that affects the successfulness of an e-learning. Among these learner characteristics, learning style is an key indicator of how a student learns and likes to learn, and how an instructor teaches to successfully address the needs of the individual students (Chang, Kao, Chu, & Chiu, 2009; Tseng, Chu, Hwang, & Tsai, 2008).

Learning style is a distinctive and habitual manner of acquiring knowledge, skills or attitudes through study or experience while learning preference is favoring of one particular mode of teaching over another (Marković, & Jovanović, 2011). There are a lot of learning style models developed in past fifty years. Witkin, Oltman, Raskin & Karp (1971) first systematically used a Group Embedded Figures Test to identify field independence of the learner. Kolb (1984) employed The Learning Style Inventory as the instrument to classify learner into four categories as convergent learners, divergent learners, assimilators, and accommodators. Keefe (1987) developed a learning style test. It can identify learners into four skill categories: Sequential Processing Skill, Discrimination Skill, Analytic Skill and Spatial Skill. Felder & Silverman’s (1988) model, however, comprises the category of intuitive/sensitive, global/sequential, visual/verbal, inductive/deductive and active/reflective, which can be used to discriminate 32 learning styles. Finally, Fleming defines learning style as “an individual’s characteristics and preferred ways of gathering, organizing, and thinking about information. VARK is in the category of instructional preference because it deals with perceptual modes (Marković, & Jovanović, 2011). The acronym VARK stands for Visual (V), Aural (A), Read/Write (R), and Kinesthetic (K).

Fleming (2012) further explained that life is multimodal. There are seldom instances where one mode is sufficient to describe complicated learner characteristics. For those who do not have a standout mode with one preference score well above other scores resulting from the VARK questionnaire, are defined as multimodal. Therefore he categorized VARK into fifteen learning style within three modes. The three modes are: single mode, dual-mode, and multimode. Fifteen learner styles are then categorized into three modes, they are: V, A, R, and K, for single mode, VA, VR, VK, AR, AK, and RK for dual-mode, finally, VAR, VAK, VRK, ARK, and VARK for multimode. Huang, Lin, and Huang (2012) criticized that several studies investigated the relationship between learning style and performance most have adopted a dichotomous definition of learning style that does not offer sufficient information for an in-depth investigation of the relationship. The VARK learning style model, however, is among the few that allow categorizes learner into bi/multi-learning style modes. This unique characteristics of VARK classification scheme is particular suitable for the present study.

3. Method

3.1 Research Goal and Questions

The goal of the study was to examine whether a comprehensive AR-facilitate learning process, including lecture, hands-on AR experience, peer discussions, and written exercises, would have extensive adaptions of different learner styles. Based on the assumption that there will be less effects of learning styles on learners achievements while AR-facilitate learning process applied, under the VARK learning style classification scheme, the following research questions were issued:

1. Is there a significant effect of learning style, in terms of single, dual, or multi- mode, on learning achievement while AR-facilitate learning process applied?
2. Is there significant differences among the four single-mode learning styles (V, A, R, and K) on learning achievement while AR-facilitate learning process applied?
3. Is there significant differences among the six dual-mode learning styles (VA, VR, VK, AR, AK, and RK) on learning achievement while AR-facilitate learning process applied?
4. Is there significant differences among the four multi-mode learning styles (VAR, VAK, ARK, and VARK) on learning achievement while AR-facilitate learning process applied?

3.2 The AR Learning Kit

The AR learning kit used for the experiment consists of three components: a sun/earth relation turntable, a computer with screen, and a webcam that captures the birds-eye-view of the turntable. This AR learning kit is able to display a day/night sensitive map and a schematic diagram simulating the changes of pole shadows. The trigger image is the earth on a turntable that simulates the revolution of earth around the sun. There are three images synchronously display on a computer screen. Image on top of the screen displays the overlay images of the earth and the sun. Image on the bottom left side is the day/night sensitive map and image on the bottom right side is the pole shadow schematic diagram. Students are able to turn the earth on the turntable manually and three images will simultaneously simulate the situations in accordance with the date and time displayed on the most top of the screen. Please refer to Figure 1 for the actual orientation of the AR learning kit.

![Figure 1. The Orientation of AR Learning Kit.](image)

3.3 Lesson Plan

As AR learning were found to be particular helpful for learning spatial concepts, earth science phenomena: “Day, night, and seasons” that involved spatial orientations of was selected as the learning content. Specific learning objectives were designed as: 1) to understand the interchange of day and night is affected by the earth rotation; 2) to understand the alternation of four seasons is affected by the revolution of the earth; 3) to understand the position and the length of pole shadows are affected by the rotation and revolution of the earth, respectively. Anderson and Krathwohl’s (2001) revision of Bloom's taxonomy of educational objectives was employed as the guideline for constructing the instructional content. Knowledge, understanding, application, and analysis were taken as the four dimensions for instructional content. In knowledge dimension, the phenomena of rotation and revolution of the earth, the interchange of day and night, as well as the move of pole shadow are described and demonstrated. In understanding dimension, the reasons why the phenomena happen are described. In application dimension, students are required to operate the AR learning kit, observe the relationship between the move on the turntable and the coordinate change on AR displays on screen. Finally, in analysis dimension, exercises are provided to allow students use acquired knowledge to analyze given situations and resolve problems.

A three-stage AR instructional process was designed to bring about the learning objectives. The three stages are: gaining attention stage, learning with AR stage, and summarization stage. In the gaining attention stage, the instructor gave lecture on these natural phenomena and showed examples of the relationships between seasons and length of a day as well as the changes of the pole shadows. In the learning with AR stage, students were divided into groups by five, each group operated the AR-learning kit (as shown on graph 1 below). An assistant was assigned to each group to help the operation. These assistants also raised questions after hands-on AR experience for initializing within group peer discussions. Finally, in the summarization stage, students were obtained reinforcements of learned
concept. A reflective sheet that required students to answer fill-in-blank questions by operating the AR learning kit was given. A summary of lesson plan for the experiment, including learning activities, amount of time spent, and the learning materials for each stage are shown on table 1.

Table 1: Summary of the Lesson Plan.

<table>
<thead>
<tr>
<th>Stage</th>
<th>Learning activities</th>
<th>Time</th>
<th>Material</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gaining attention</td>
<td>Instructor gives lecture and shows examples,</td>
<td>25 min.</td>
<td>PowerPoint briefings</td>
</tr>
<tr>
<td>Learning with AR</td>
<td>Students operate the AR-learning kit in groups,</td>
<td>50 min.</td>
<td>AR learning kit</td>
</tr>
<tr>
<td></td>
<td>Peer discussions</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Summarization</td>
<td>Students obtain reinforcement of learned concepts</td>
<td>15 min.</td>
<td>Reflective sheet (written form)</td>
</tr>
</tbody>
</table>

3.4 The Instrument

Fleming’s VARK learning style questionnaire employed to classify the multiple-tendency of learning styles [www.business.vark-learn.com]. The Younger Version revised on September, 2007 was used to fit the age range (13-15 years old) of our subjects. They are totally 16 questions in the questionnaire. Four selections of possible answers are available for each question. Each answer refers to one of the VARK category. Multiple selections of these answers are allowed. Totally 15 categories can be classified by this questionnaire. The reliability of the Chinese version of the questionnaire is Cronbach’s α= .83.

A pretest and a posttest for evaluating the learning achievement were constructed. The revised taxonomy aforementioned was taken as the guideline for constructing test items. Both pretest and posttest have 18 single-answer multiple choice questions. The knowledge and understanding dimensions have four questions, and application and analysis dimensions have 5 questions. The pretest and posttest are designed as parallel forms. The reliability of the pretest and posttest are α= .74 and .84, respectively.

3.5 The Experiment

The experiment was done in two separate junior high schools in New Taipei City of Taiwan in a period of three months. Totally 144 students in five seven-grade classes were selected as subjects. One instructor and seven teaching assistances were involved in the instructional process. An independent three-stage AR instructional process was performed for each of the five classes. A pretest was given before, and a posttest was given after each instructional process was performed. Several students were randomly asked for a brief interview to understand student’s interests and motivations on using the AR learning kit.

4. Results and Discussions

4.1 The Distribution of Learning Styles

According to Fleming’s VARK classification scheme, students were categorized into 15 learning styles within single, dual, and multiple learning modes. The most prevalent learning style type was “VARK”. 35 out of 144 students fell into this category. “V” and “VK” were the two categories that had least students fit in. In terms of learning style mode, most students had multi-modal learning style (80), and least students were single-modal (26). Detailed information please refer to Table 2.

Table 2: Distribution of Learning Mode and Learning Style.

<table>
<thead>
<tr>
<th>Learning Mode</th>
<th>Learning Style</th>
<th>Number of Samples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Single-modal</td>
<td>V</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>A</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>4</td>
</tr>
</tbody>
</table>

207
The effects of the three modes of learning style: single, dual, and multiple modes on learning achievement were statistically analyzed. An ANCOVA was performed to examine the significance of the mean differences among the three modes. Pretest score was used as the covariance. The result indicated that there was no significant difference on learning achievement found among these three modes ($F_{2,140} = .017$, $p=.983$). As we also found that there is a significant pretest-posttest gains ($t_{286}=10.346$, $p<.001$), it was evident that the three-stage AR-facilitate learning process was adaptive for learners with any learning mode. Table 2 summarizes the ANCOVA.

Table 3: Summary of ANCOVA for Learning Style Mode.

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>$F$</th>
<th>$\eta^2$</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mode</td>
<td>.327</td>
<td>2</td>
<td>.163</td>
<td>.017</td>
<td>.000</td>
<td>.983</td>
</tr>
<tr>
<td>Error</td>
<td>1336.534</td>
<td>140</td>
<td>9.547</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

4.3 The Effects of Learning Styles On Achievement

We further examined the effects of learning style types on student’s learning achievement. Three separate ANCOVA was performed for the three learning style modes.

There are four types of single-modal learning styles: “V”, “A”, “R”, and “K”. The result of ANCONA indicated a near significant result ($F_{3,21} =2.581$, $p=.081$). In order to avoid possible type II error, LSD post hoc comparisons was done. The result of the post hoc analyses indicated that the achievement for “R” type learners was significantly better than “A” type ($p=.030$) and “K” type ($p=.012$). This result seems to be incoherent with Fleming’s account that “R” type of learner prefers to use text-based materials. The possible reason is that the summarization stage helped more for “R” type of students in reviewing the concepts learned during the AR operation. Table 3 summarizes the ANCOVA and the LSD comparisons for the single-modal learning styles.

Table 4: Summary of ANCOVA and Post Hoc For Single-Modal Styles.

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>$F$</th>
<th>$\eta^2$</th>
<th>$p$</th>
<th>Post hoc</th>
</tr>
</thead>
<tbody>
<tr>
<td>Learning style</td>
<td>3</td>
<td>2.581</td>
<td>.269</td>
<td>.081</td>
<td>R &gt; A, R &gt; K</td>
</tr>
<tr>
<td>Error</td>
<td>21</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The differences among six types of dual-modal learning styles: “VA”, “VR”, “VK”, “AR”, “AK”, and “RK”. The result of ANCOVA showed a non-significant result ($F_{5,31}=1.000$, $p=.434$), there was no differences on learning achievement regarding dual-modal learning styles. Table 4 summarizes the ANCOVA for the dual-modal learning styles.
Finally, we examined the effects of multi-modal learning styles. Again, the result of ANCOVA exhibited that there was no significant difference among “VAR”, “VAK”, “VRK”, “ARK”, and “VARK” types of learners. Table 5 summarizes the ANCOVA for the multi-modal learning styles.

Table 5: Summary of ANCOVA for Dual-Modal Styles.

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>$\eta^2$</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Style</td>
<td>44.383</td>
<td>5</td>
<td>8.877</td>
<td>1.000</td>
<td>0.139</td>
<td>0.434</td>
</tr>
<tr>
<td>Error</td>
<td>275.140</td>
<td>31</td>
<td>8.875</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Finally, we examined the effects of multi-modal learning styles. Again, the result of ANCOVA exhibited that there was no significant difference among “VAR”, “VAK”, “VRK”, “ARK”, and “VARK” types of learners. Table 5 summarizes the ANCOVA for the multi-modal learning styles.

Table 6: Summary of ANCOVA for Dual-Modal Styles.

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>$\eta^2$</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Style</td>
<td>85.252</td>
<td>4</td>
<td>21.313</td>
<td>1.858</td>
<td>0.074</td>
<td>0.127</td>
</tr>
<tr>
<td>Error</td>
<td>860.298</td>
<td>75</td>
<td>11.471</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

5. Conclusion

An adaptive education combines the development of an individual’s initial competence with alternative environments matched to different styles of learning. The adaptation of individual differences to e-learning has been discussed for decades. However, most efforts have been made to develop intelligent programs to select appropriate instructional paths and/or to determine the amount of instruction to be given based on individual learner’s on-task performance. To avoid complicated adaptive computer algorithms, we developed a comprehensive AR-facilitate e-learning process that is able to take place in regular classroom settings. Instructional activities included life lectures, student manipulations of virtual objects, peer discussions, and written exercises. An experiment employing the AR process was done with an earth science learning unit. The VARK learning style classification scheme was used. The results showed there was no significant difference in learning achievement of students with different mode of learning styles.

This result is promising. The present study provides an alternative rationale for developing adaptive e-learning without involving complicated adaptive algorithm. A comprehensive AR-facilitate e-learning process could make the regular classroom to be more adaptive to students with different learning styles.

Although findings of this study are potentially supporting the development of an alternative ICT-based e-learning strategy, some inherent limitations must be addressed. As we admitted the non-significant results related to the differences among learning styles, there were some risks of gaining type II errors. Although most of the non-significant decisions were made on a reasonably reliable basis ($p > .3$), sparse significant results were found in post hoc analyses of single-modal learning styles ($R > A$, $R > K$). Another limitation of the study is that the generalizability of the research findings is restrained because only a single learning unit within a single subject matter was implemented for the experiment. For future studies, we suggested that larger sample sizes and extensive subject matters need to be concerned.

Acknowledgements

Funding of this research work is supported in part by the Ministry of Science and Technology of Taiwan (under research numbers NSC 102-2511-S-003-023- and MOST 103-2420-H-003-031-).

References


The Effects of AR-based Instruction on Students’ Learning Performance, Motivation and Self-efficacy in Programming Learning

Gloria Yi-Ming KAO* & Cheng-An RUAN
Graduate Institute of Digital Learning and Education, National Taiwan University of Science and Technology, Taiwan
*gloriakao@gmail.com

Abstract: Recent studies have reported benefits of Augmented Reality (AR)-based instruction in various learning domains. However, few studies were done to explore its effects in programming learning. In this study, we devised an AR-based instruction with high (puzzle cards) or low (fixed card) interaction levels to assist programming learning. The results showed that students’ learning performance and self-efficacy were improved after the experiment, which indicated AR-based learning did have positive effects though no difference between high and low interaction could be determined at this stage.

Keywords: augmented reality, programming learning, learning performance, motivation, self-efficacy

1. Introduction

1.1 Augmented Reality (AR)

Augmented Reality (AR) is a technique that adds virtual objects in a real environment, which can redeem the lack of information in a real environment (El Sayed, Zayed, & Sharawy, 2010). Physical and virtual tools are commonly utilized as assistant teaching tools in classrooms; however, it is hard to combine the strengths of the two tools. After AR was invented, which integrated physical and virtual functions, learners were supplied with better learning experiences (Bujak et al., 2013). AR enables learners get close to real environments from learning environments and supplies richer sensory experiences. Also, it makes learners have the opportunities to operate physical objects and then to interact with virtual ones (Wojciechowski & Cellary, 2013).

1.2 Programming Learning

Programming is a big role in science, technology, engineering and mathematics (STEM) fields. Especially, to those major in computer science, programming course plays an important role. Nevertheless, learning programming is difficult to the beginners who are not in the related fields (McCracken et al., 2001). Despite of learners’ ages, programming is always difficult for beginners (Kelleher & Pausch, 2005). In Taiwan, traditional way to learn programming usually use textbooks or run sample program codes with computers, and then observe the execution results. However, beginners are not able to realize the programming process or the results. This can lead to poor learning motivation or performance.

1.3 AR in Education

Recently, many studies have shown that AR-based learning has positive influence on students’ learning, such as learning performance (Lin, Duh, Li, Wang, & Tsai, 2013), learning motivation (Di Serio, Ibáñez, & Kloos, 2013) and self-efficacy (Kamarainen et al., 2013).
Therefore, AR is commonly used as an educational tool. Liu and Tsai (2013) employed teaching materials combined with AR in an English writing course, trying to reduce the difficulties of the students while learning second language. Results indicated that learners under this situation could construct contents and knowledge much easier; moreover, more meaningful articles were produced, and language learning performance was then increased. Chang, Chang, Sung, Chao and Lee (2014) developed an AR guide system and then used it in an art appreciation course. Compared to general audio guide and non-guide environment, learners in AR guide system group had more fluent experience and better learning performance. Chang, Wu and Hsu (2013) stimulated Fukushima nuclear disaster and explored the situation of nuclear pollution by using AR technique. The result show that AR based environment could improve students’ comprehension and increase their sensorial immersion. Ibáñez, Di Serio, Villarán and Delgado Kloos (2014) implemented AR in a basic course of electromagnetism. The results showed that, in this environment, students could not only understand the phenomena and concepts of electromagnetism more efficiently but reach higher flow experience levels, compared to web-based learning environment. AR is used in many different subjects, yet cases of programming course are seldom found. Consequently, this study explores how AR-based learning influence students’ learning performance and motivation in a programming course.

2. System design

2.1 Software Development

Aurasma, a cross-platform AR development system developed by Hewlett-Packard Development Company in 2011, supplies simple operating interface to make it easier for developers to produce AR contents. Aurasma enables users to connect to the database to get the latest AR contents by using the mobile device.

Based on the Aurasma, we devolved an AR-based application with high (puzzle cards) or low (fixed card) interaction levels to assist programming learning. Students could freely assemble puzzle cards to observe the corresponding results. With high interaction level, students are encouraged to try and figure out what combinations could work or not. With low interaction level, students could only watch default animations with fixed cards. We used pictures from the Scratch program, developed by Massachusetts Institute of Technology in 2003, in our design.

2.2 Operating process

Learners followed the teacher instruction, assembled different puzzle cards (as Figure 1 shows), and used the Aurasma in mobile device to observe and compare the differences among various combinations of puzzle cards. While the puzzle cards are assembled correctly, the programming operating animation will be displayed (as Figure 2 shows). However, the animation will not be displayed when the assembling is not correct.

![Figure 1. Assemble different puzzles.](image-url)
2.3 Learning Content

The learning content of the study was three main structures of flowchart (as Figure 3 shows) in computer programming. They were 1) Sequence structure, which followed certain order, operating the description separately; 2) Selection structure, which operates the program according to the conditional determination; 3) Iteration structure, which operates the descriptions repeatedly, until the descriptions match the breaking condition, the operation ends. Most Algorithms can be consisted of these three structures. This course is not only an important role but an essential part in programming courses. Several combinations of puzzles were supplied. Puzzles can be assembled with two combinations in Sequence and selection structures, but they can be assembled with three combinations in Iteration structure.

![Figure 3. Three main structures of flowchart.](image)

3. Experiment design

3.1 Participants

Sixty-five seventh graders from two mixed-ability classes in a junior high school in North Taiwan participated in the experiment, the mean age was 13. Thirty-one (15 females and 16 males) were in the experimental group, while thirty-four students (16 females and 18 males) were in the control group. High interactive AR-based learning systems (puzzle cards) were used in the experimental group, while low interactive AR-based learning system (fixed cards) was in the control group.

Mobile devices were not used in previous teaching process, so the students were instructed to operate the mobile devices and the AR system first before the experiment was carried out. In both of the two groups, there were 7 students had learned Scratch, a software which can program interactive
animations. Their average learning time was one year. Therefore, a pretest was adopted to exclude the differences among the students caused by prior knowledge.

3.2 Learning activities and environment

In order to fit in the teacher’s original teaching style, the experiments were taken place in a computer classroom. During the experiment, computers were only used for broadcasting PowerPoint slides for instructional purpose. For both the experimental and control groups, each two students were equipped with one mobile device (iPad) and were allowed to discuss with each other.

As Figure 4 shows, high interactive AR-based learning systems were used in the experimental group. This group was given puzzle cards and was asked to freely assemble those puzzles to yield and observe different program execution outcomes. While students in the control group with low interactive AR-based learning system could only get fixed cards and observe the default program execution results.

![Figure 4. Experiment group with puzzle cards and control group with fixed cards.](image)

3.3 Experiment procedure

Figure 5 shows the four stages of this study and describe as below:

- **Stage 1:** A pretest was adopted to examine the students’ prior knowledge about programming; pre-questionnaires were employed to probe students’ learning motivation and self-efficacy. Students were firstly explained the ways to fill out the sheets and were not allowed to discuss during the test. This stage took ten minutes.

- **Stage 2:** With one mobile device equipped, two students grouped and were taught how to operate the mobile devices and the learning system. This stage took ten minutes. The researcher checked the students’ status to make sure they were able to operate the tools.

- **Stage 3:** Programming course was started. There were four main units in this course. Students observed and learned with mobile devices after the teacher’s instruction. This stage took sixty minutes.

- **Stage 4:** A learning performance posttest, a learning motivation post-questionnaire and self-efficacy post-questionnaire were adopted. Students were not allowed to discuss during the test. This stage took twenty minutes.
4. Analysis methods and results

4.1 Analysis tools

Learning performance test: A self-edited performance test sheet was used, and the expert validity was constructed after the scale was revised by three computer and information science teachers. A pretest was employed to examine the students’ prior knowledge about programming. Each of the ten matching questions in the sheet was ten points. A post-test was adopted to explore students’ knowledge and applied abilities about programming. There were five matching questions and five applicant questions, and each question was ten points. From this test, we wanted to understand if the learning performance was improved after the experiment.

Learning motivation and self-efficacy: The learning motivation questionnaire adopted in the study was a revision from Hwang, Yang and Wang (2013) and included seven questions. The self-efficacy questionnaire was a revision from Wang and Hwang (2012) which included eight questions. Both the questionnaire adopts five-point Liker rating scheme, “1” means very disagree while “5” means very agree. All the descriptions in the questionnaire were positive (e.g. I think this course is meaningful and worthy learning, I believe that I can get great score in the assignments). For the purpose to explore the difference before and after the intervention, the two questionnaires were employed both before and after the experiment to investigate students’ perceptions toward learning motivation and self-efficacy. The original Cronbach’s alpha of the motivation questionnaire was 0.79, while the revised one was 0.883. The original Cronbach’s alpha of the self-efficacy questionnaire was 0.916, while the revised one was 0.919.

4.2 Method

The present study analyzed and processed the data by using statistical software. Firstly, the descriptive statistic, mean and standard deviation, would show the differences among learning performance, learning motivation and self-efficacy. Secondly, after pre-test and pre-questionnaire, independent t-Test will be employed to examine if there are any differences between the two groups. Thirdly, ANCOVA will be used to explore if there are any differences in post-test and post-questionnaire between the two groups. Above three steps will be adopted to observe how high and low interactions influence AR-based instruction. Lastly, how AR-based instruction influence programming learning will be discussed. Also,
dependent t-Test will be used to examine whether there are any significant differences between pre-and-post test, and pre-and-post questionnaire.

4.3 Results

Several independent t-Tests were conducted to examine the difference between experimental and control group before the experiment. As shown in Table 1, there is no significant difference between experimental and control group (pre-test: \( t=-0.403, p>0.05 \), pre-motivation: \( t=1.103, p>0.05 \), pre-self-efficacy: \( t=0.136, p>0.05 \)) which means before the learning activity, students in the two groups have equivalent prior knowledge, learning motivation and self-efficacy.

**Table 1: Independent t-Test result of Pre-test, Pre-motivation and Pre-self-efficacy of Experimental group and Control group.**

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>Mean</th>
<th>SD</th>
<th>t</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Pre-test</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Experimental</td>
<td>31</td>
<td>47.097</td>
<td>19.008</td>
<td>-0.403</td>
<td>0.689</td>
</tr>
<tr>
<td>Control</td>
<td>34</td>
<td>48.824</td>
<td>15.524</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Pre-motivation</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Experimental</td>
<td>31</td>
<td>4.143</td>
<td>0.655</td>
<td>1.103</td>
<td>0.274</td>
</tr>
<tr>
<td>Control</td>
<td>34</td>
<td>3.975</td>
<td>0.574</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Pre-self-efficacy</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Experimental</td>
<td>31</td>
<td>3.347</td>
<td>0.948</td>
<td>0.136</td>
<td>0.892</td>
</tr>
<tr>
<td>Control</td>
<td>34</td>
<td>3.320</td>
<td>0.582</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

After the learning activity, several analysis of covariance (ANCOVA) were used to evaluate the difference between experimental and control group in terms of learning performance, motivation or self-efficacy. The pre-test score of each evaluation was used as covariate while the post-test score of each evaluation was used as dependent variable. Table 2 shows that the results are not significant (post-test: \( F=0.082, p>0.05 \), post-motivation: \( F=1.216, p>0.05 \), post-self-efficacy: \( F=0.142, p>0.05 \)) which means high or low interactive AR-based leaning would not impact on students’ learning performance, motivation or self-efficacy.

**Table 2: ANCOVA result of Post-test, Post-motivation and Post-self-efficacy of Experimental group and Control group.**

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>Mean</th>
<th>SD</th>
<th>F</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Post-test</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Experimental</td>
<td>31</td>
<td>69.839</td>
<td>19.701</td>
<td>0.082</td>
<td>0.775</td>
</tr>
<tr>
<td>Control</td>
<td>34</td>
<td>68.618</td>
<td>17.132</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Post-motivation</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Experimental</td>
<td>31</td>
<td>4.014</td>
<td>0.654</td>
<td>1.216</td>
<td>0.274</td>
</tr>
<tr>
<td>Control</td>
<td>34</td>
<td>3.870</td>
<td>0.554</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Post-self-efficacy</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Experimental</td>
<td>31</td>
<td>3.581</td>
<td>0.935</td>
<td>0.142</td>
<td>0.707</td>
</tr>
<tr>
<td>Control</td>
<td>34</td>
<td>3.522</td>
<td>0.607</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Several dependent t-Tests were conducted on exploring how AR-based learning influences students’ learning performance, motivation and self-efficacy, as shown in Table 3. There is a significant difference between learning performance (\( t=-8.634, p<0.05 \)) and self-efficacy (\( t=-2.557, p<0.05 \)), yet learning motivation is not significantly different (\( t=1.559, p>0.05 \)). The results showed overall after the learning activity, students’ learning performance and self-efficacy were increased significantly, although no difference between the puzzle cards and fixed cards could be determined.

**Table 3: Dependent t-Test result of Pre-test, Post-test, Pre-motivation, Post-motivation, Pre-self-efficacy and Post-self-efficacy.**

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>Mean</th>
<th>SD</th>
<th>t</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre-test</td>
<td>65</td>
<td>48.000</td>
<td>17.157</td>
<td>-8.634***</td>
</tr>
<tr>
<td>Post-test</td>
<td>65</td>
<td>69.200</td>
<td>18.266</td>
<td></td>
</tr>
<tr>
<td>Pre-motivation</td>
<td>65</td>
<td>4.055</td>
<td>0.615</td>
<td>1.559</td>
</tr>
<tr>
<td>Post-motivation</td>
<td>65</td>
<td>3.938</td>
<td>0.603</td>
<td></td>
</tr>
</tbody>
</table>
5. Conclusions and discussion

The present study explores how high and low interactive AR-based learning influence students’ learning. As the results showed, different interactive levels with puzzle or fixed cards did not impact on students’ learning performance, motivation and self-efficacy. Instead, overall students’ learning performance and self-efficacy were improved after the experiment, which indicated the innovative use of AR into programming learning did have positive effects although no difference between high and low interaction could be determined at this stage.

Convenience sampling was conducted in this study, which was a limitation of the study. The results can only represent the learning performance of the students in the experiment, but cannot infer the overall students’ in other areas. Further, time duration was another limitation of the study. Owing to the time, the states of students’ learning and how much did they exactly learned were not thoroughly considered.

Future studies are suggested dividing the participant groups into high-interactive AR, low-interactive AR, and traditional learning to further investigate how AR-based learning influences students’ learning.

Acknowledgements

This research was funded by the National Science Council of the Republic of China, Taiwan, under contract number NSC 102-2511-S-011 -005 -MY3.

References


Implementation of Student-associated Game-based Open Inquiry in Chemistry Education: Results on Students' Perception and Motivation

Keeratika MEESUK\textsuperscript{a}, Niwat SRISAWASDI\textsuperscript{b}\textsuperscript{*}

\textsuperscript{a}Science and Technology Education Program, Faculty of Education, Khon Kaen University, Khon Kaen, Thailand
\textsuperscript{b}Division of Science, Mathematics, and Technology Education, Faculty of Education, Khon Kaen University, Khon Kaen, Thailand
\textsuperscript{*}niwsri@kku.ac.th

Abstract: Educational computer game refers to the use of digital technology to promote learning performance and experience through game-based activity. Currently, researchers mentioned that digital game-based learning can promote students’ interest and enhance learning outcomes. As such, this study aims to develop inquiry-based learning process with the support of digital game for chemistry learning about ionization energy. This paper reports research findings on two studies. First, a digital game regarding chemistry concept of ionization energy has been created and 29 twelfth-grade students participated to play the game. The students’ perceptions towards the game were measured and results showed that the game have significant effect in fostering their perceptions. Another, 87 tenth-grade students were recruited in this study, and they were divided into an experimental group (N=43) and a control group (N=44). The experimental group participated with student-associated game-based open inquiry, called SA GOI, class and another were assigned to conventional class. The results indicated that students in SA GOI class have changed science motivation over the SA GOI learning experience. This could be implied that the student-associated game-based open inquiry could be an alternative way for promoting chemistry learning in schoolscience.

Keywords: Digital game, inquiry-based learning, ionization energy, science motivation

1. Introduction

Edutainment is an educational concept which aims to make a combination between education and entertainment. As such concept, educational computer game could be considered as an alternative form of edutainment that learners can learn lessons of subject matter by playing the game. By the way, researchers mentioned that educational game is different from other edutainment by its nature in requiring strategies, proving hypothesis, solving problems, and it usually use higher-order thinking rather than memorization. The main characteristics of educational game are the challenging to achieve the objectives, and providing a specific situation and reward for engaging and motivating learners which act as the players (Prensky, 2001; Papastergiou, 2009). A recent study reported about how digital game support learners’ motivational and cognitive processing. Huang (2011) indicated that learners have more confident in learning after playing with educational game. In an addition, using game in education increased students' perceived learning, enjoyment and flow of learning experience (Barzilai and Blau, 2014).

For chemistry class, not many study investigated influence of educational game on students' learning outcomes. In fact, chemistry knowledge is often abstract, complex, and complicate in representations of the chemistry knowledge. The use of digital game may be help student increasing learning interest in chemistry, motivation, and attitude towards chemistry learning. Therefore, this
study aims to create educational computer game as an inquiry tool to learn chemistry in concepts of ionization energy.

2. Research Questions

Based on the above mentioned rationale, the goals of this study was a couple: (i) to investigate students’ perceptions towards an educational computer game, called "The IE War"; and (ii) to investigate students' science motivation delivered in a proposed open-inquiry learning process with support by an educational computer game, called "Student-associated Game-based Open Inquiry (SAGOI)”. Specifically, the following questions were answered:

- Do the students interacted with "The IE War" perform significantly better by perceptual constructs i.e. perceived learning, flow, enjoyment, and satisfaction?
- Do the students engaged in "SAGOI" perform significantly better by motivational constructs towards science learning i.e. intrinsic motivation, self-determination, self-efficacy, career motivation, and grade motivation?

3. Backgrounds and Research Hypotheses

2.1 Educational Computer Game and Students' Perceptions

2.1.1 Perceived Learning

Perceived learning relates to a retrospective evaluation of the learning experience and can be defined as a set of beliefs and feelings one has regarding the learning that has occurred (Caspi and Blau, 2011). The perceived learning is about the new information was obtained and person can get the new understanding, subjective evaluation of learning by learners themselves. Researchers mentioned that perceived learning is connected to emotion as flow, enjoyable, and satisfaction (Chu and Hwang, 2010). Regarding in context of educational computer game, when learners are immersed in game-based learning environment, they can judge themselves in the learning process and quality of how to get the knowledge from game, so game can help learned and practiced (Giannakos, 2013). According to the abovementioned, a research hypothesis was set to examine in this study as following.

H1. Students’ perceived learning between pre- and post-interacting with educational computer game has development.

2.1.2 Flow

Flow is a state of deep concentration in which thoughts, intentions, feelings, and all of the senses are focused on the same goal (Csikszentmihalyi, 1990; Barzilai and Blau, 2014). The experience of flow would happen when person who take part in challenge situations or activities that need skills. Flow depends on a chance to concentrate, an immediate feedback, a sense of control, and a clarify goal (Barzilai and Blau, 2014). As such, if learners concentrate with the learning experience of educational computer game, the flow of learning would occur during playing the game. According to the abovementioned, a research hypothesis was set to examine presented as following.

H2. Students flow between pre- and post-interacting with educational computer game has development.

2.1.3 Enjoyment

Enjoyment is the condition of having and using technology, e.g. educational computer game that is good or pleasant. The enjoyment of player is a key goal, related with an easy to use of game and enjoyment was found to have valuable in explaining objective to use applications (Giannakos, 2013).
When learners which act as players of game fail to pass the game task, they would get disappointment and attempt to replay again. As such, the enjoyment can help reduce worry to learn and feel more confident when learners succeed. Accordingly, if the educational computer game can enjoy learners, then they would like to learn more and think positive to the subject. Based on this aspect, below is a research hypothesis for examining the impact of educational computer game on enjoyment.

**H3.** Students’ enjoyment between pre- and post-interacting with the educational computer game has development.

### 2.1.4 Satisfaction

Satisfaction is the individual awareness of how well a learning environment supports academic success (Lo, 2010). It is relevant to instructional method that learners can think and learn, so their satisfaction can help to get how academic success. At the moment to learn with educational computer game, if it gets positive response from learners that means they reach to positive learning experience with also. In an addition, satisfaction can yield positive of learning performance and can improve learning outcome (Giannakos, 2013). To investigate an impact of educational computer game, a hypothesis was set as the following.

**H4.** Students’ satisfaction between pre- and post-interacting with the educational computer game has development.

### 2.2 Student-associated Game-based Open Inquiry and Science Motivation

Science motivation refers to the motivation of students to learn science within their emotional which stimulate, control and support in science learning behavior. Therefore, science motivation could be achieved to learners when activate their behaviors by asking the questions, doing experiments, and collaborative learning (Schunk, Pintrich and Meece, 2008; Glynn et al., 2011). Researchers stated that science motivation consists of five motivational constructs: *intrinsic motivation*, an internal state of satisfaction to learn science because it will be good for itself; *self-determination*, the controlling of students’ belief that they have when learning science; *self-efficacy*, students can bring their belief connect and manage to achieve learning science; *career motivation*, students learn science to get a good work in the future; and *grade motivation*, learning science to have a good score (Glynn et al., 2011). The following research hypothesis was another one which the researchers expected to examine in this study.

**H5.** Students’ science motivation between pre- and post-participating in the student-associated game-based open inquiry has development.

### 4. Research Design

This study used two different research designs: one-group pretest-posttest design to examine impacts of the proposed digital game, The IE War, on students’ perception and a simple two-group comparison to examine the effects of an instructional intervention, The SAGOI, on students’ science motivation, as study 1 and 2 respectively.

### 5. Study 1

#### 5.1 Participants

The subjects of this study were 29 twelfth-grade students in a public school at the northeastern region of Thailand. The age range of the students was 17-18 years, and all of them were females. They were attending a chemistry course for basic education level and all of them have satisfactory skills on basic
computer and information and communication technology, but they have no experience yet using educational computer game in chemistry learning.

5.2 Learning Materials

The digital game on chemistry of ionization energy was designed to comprise three stages and ten levels of playing. The IE War is the first stage and it was created in style of shooting game. The goal of this educational game is to facilitate student getting the definition of ionization energy. To complete The IE War, there were five playing levels as followings. Figure 1 illustrates a flow chart of The IE War game.

![Flow chart of The IE War game](image)

**Figure 1.** Flow chart of The IE War game.

**Level 1:** The players have only two bullets, which are IE and IE++ bullet and various gaseous atoms fall down to the ground. The players are able to shoot the atoms by the IE bullet, and then the atom changes its state to be gaseous ion if its energy is corrected. For one round, the players have only three hearts power and they must play to get 10 points for passing any level. Finally, the players who can pass this level would get another bullet, a sublimating bullet.

**Level 2:** The players have three bullets, which are IE, IE++, and sublimating bullet. In this level, various solid atoms fall down to the ground. The players must choose the sublimating bullet to sublime solid into a gas, before using IE and IE++ respectively. In the end of this level, the players who complete this level would get the evaporating bullet for the next level.
Level 3: The players have four bullets, which are IE, IE++, sublimating, and evaporating bullet. Likewise, various liquid atoms fall down to the ground and the players must use the evaporating bullet to shot the liquid into a gas, before using IE and IE++. Finally, the players who complete this level would get another bullet called break bond.

Level 4: The players have five bullets, which are all previous four and the break bond bullet. Gaseous molecules fall down and the player must shot them by the break bond bullet to break the gaseous molecules into gas, before using IE and IE++ in next.

Level 5: In this level, all state of matter including gas, solid, liquid, and gaseous molecules fall down to the ground. The players have all five bullets, and must think how to use the bullet correctly and then if the points are high enough, the players would finish this game.

5.3 Instrument

An 18-item perception questionnaire was used to measure students’ perception on four subscales: perceived learning, flow, enjoyment, and satisfaction. All of these 5-point Likert scale items obtained from Chu and Hwang (2010) and Barzilai and Blau (2014). From the English version, an identical version in Thai was constructed and two experts were recruited to identify communication validity. For the Thai version, reliability for the overall items was 0.902. There were four items on perceived learning subscale and its reliability was 0.754. The five items on flow subscale indicated that its reliability was 0.661. For enjoyment and satisfaction subscales, the reliability for them was 0.751 and 0.857 respectively.

5.4 Data Collection and Analysis

The participants were asked to complete the perception questionnaire to measure their pre-perceptions on perceived learning, flow, enjoyment, and satisfaction for 10 minutes. After completing the instrument, they were exposed to play The IE War game for 20 minutes. After finishing the game, the students’ post-perceptions were examined by the same questionnaire for 10 minutes. To compare the pre-post perceptions on each subscale, paired t-test in SPSS was used to indicate the difference.

5.5 Results and Discussion

Compared to pre-perception scores, the results of post-perception indicated significant higher of students’ perception scores on perceived learning (t = 3.324, p < .01), flow (t = 5.158, p < .01), enjoyment (t = 2.480, p < .01), and satisfaction (t = 4.297, p < .01). This indicated that students have increased their positive perception towards playing the game. Figure 2 illustrates a graphical representation on students’ pre- and post-perception scores. The figure showed that the IE War game can affect students’ perceived learning, flow, enjoyment, and satisfaction to learn science effectively.

![Figure 2. Compare mean scores between pre- and post-questionnaire of four scales.](image)
6. Study 2

6.1 Participants

A total of 81 tenth-grade students were recruited into this study. They were divided into an experimental group (N=43) and a control group (N=44). The experimental group participated with SAGOI class and another was assigned to conventional class as a control group. They were attending a chemistry course for basic education level and all of them have satisfactory skills on basic computer and information and communication technology, but they have no experience yet using educational computer game in chemistry learning before.

6.2 Learning Materials and Activity

The IE War game and another two game related factors that affect to ionization energy were incorporated as instructional tool into open-inquiry learning process, called SAGOI. In this learning process, students will collaborative work together in small groups of three to five members. This pedagogy begins with an open-ended driving question targeted to alternative conceptions about ionization energy commonly found in students. To assist the process of hypothesis generation addressed the question, essential scientific backgrounds are provided to students. Then, students are required to perform generating testable hypotheses, designing an investigation with the games. During playing the game, each group was assigned to access Google Drive spreadsheet, preparing by instructor, for recording scores and what they found into a predetermined table. In an addition, each group was assigned to analyze the recorded data by comparing individual score and also use Google Chat for discussing in the group. When they finished the game, all groups have to communicate findings among groups by creating a PowerPoint presentation via Google Drive presentation. Finally, instructor induces students into a forum for drawing a conclusion based on evidence and collaborative explaining the result of hypotheses testing.

6.3 Instrument

A 25-item science motivation questionnaire was used to measure students' motivation to learn science on five subscales: intrinsic motivation (IM), self-determination (SDT), self-efficacy (SEC), career motivation (CM), and grade motivation (GM) (Glynn et al., 2011). From 25 items English version, the translation an identical version in Thai was constructed and Cronbach’s alpha of Thai version were 0.79, 0.81, 0.89, 0.81 and 0.85 for IM, SDT, SEC, CM and GM respectively (Srisawasdi, submitted).

6.4 Data Collection and Analysis

Students were investigated science motivation by the 5-point Likert-scale questionnaire before giving SAGOI intervention for 10 minutes. In the SAGOI class, students participated biology learning of ionization energy using the developed digital game for 230 minutes. After the instruction, students were administered by the same questionnaire again as post-test. The statistical data techniques selected for analyzing students' science motivation was repeated-measures MANOVA in SPSS.

6.5 Results and Discussion

A repeated-measures MANOVA was conducted to determine students’ science motivation scores on the five subscales. The assumption of homogeneity of variance-covariance was tested with Box’s M Test which was not significant and indicated that homogeneity of variance-covariance was fulfilled (p = .313). The results for the repeated-measures MANOVA indicated significant main effect for group (Wilks’ lambda =0.713, F (5, 75) =6.029, p<0.001, η²=0.287), and time [Wilks’ lambda=0.663, F (5, 75) = 7.634, p<0.001, η²=0.337], but interaction effect of time and group were not significant. Thus, these results indicated that students in SAGOI class have changed science motivation over the SAGOI learning experience. Univariate MANOVA on each subscale was conducted as follow-up tests to the
one-way MANOVA. The results of the univariate test for control- and experimental group students are summarized in Table 1.

Table 1: The students' subscale means of science motivation by time and univariate MANOVA.

<table>
<thead>
<tr>
<th>Subscale</th>
<th>Time</th>
<th>CG Pre-test: Mean (SD)</th>
<th>CG Post-test: Mean (SD)</th>
<th>EG Pre-test: Mean (SD)</th>
<th>EG Post-test: Mean (SD)</th>
<th>F</th>
<th>Sig.</th>
<th>η²</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intrinsic motivation</td>
<td>IM</td>
<td>17.10 (2.73)</td>
<td>17.61 (2.95)</td>
<td>17.95 (2.64)</td>
<td>19.23 (2.51)</td>
<td>8.96</td>
<td>0.004</td>
<td>0.10</td>
</tr>
<tr>
<td>Self-determination</td>
<td>SDT</td>
<td>15.39 (3.71)</td>
<td>16.54 (3.20)</td>
<td>15.65 (2.91)</td>
<td>18.00 (2.92)</td>
<td>26.27</td>
<td>0.000</td>
<td>0.25</td>
</tr>
<tr>
<td>Self-efficacy</td>
<td>SEC</td>
<td>13.17 (3.29)</td>
<td>14.56 (3.65)</td>
<td>15.83 (3.99)</td>
<td>16.98 (3.36)</td>
<td>21.11</td>
<td>0.000</td>
<td>0.21</td>
</tr>
<tr>
<td>Career motivation</td>
<td>CM</td>
<td>17.39 (3.30)</td>
<td>18.07 (3.25)</td>
<td>18.12 (2.86)</td>
<td>19.73 (3.17)</td>
<td>14.87</td>
<td>0.000</td>
<td>0.15</td>
</tr>
<tr>
<td>Grade motivation</td>
<td>GM</td>
<td>17.44 (3.38)</td>
<td>17.44 (3.67)</td>
<td>20.00 (2.70)</td>
<td>20.72 (2.26)</td>
<td>1.15</td>
<td>0.285</td>
<td>0.01</td>
</tr>
</tbody>
</table>

In Table 1, the univariate MANOVA from pre- to post-questionnaire of four subscale scores consists of IM, SDT, SEC and CM were significant differences across time. The univariate results pointed out a significant effect on IM ($F_{1,79} = 8.960, p < 0.01, \eta^2 = 0.102$), SDT ($F_{1,79} = 26.273, p < 0.001, \eta^2 = 0.250$), SEC ($F_{1,79} = 21.113, p < 0.001, \eta^2 = 0.211$) and CM ($F_{1,79} = 14.873, p < 0.001, \eta^2 = 0.158$), but another one GM, the univariate result displayed insignificant ($F_{1,79} = 0.285, p < 0.01, \eta^2 = 0.014$). The result suggested that the increase of science motivation regarding intrinsic motivation, self-determination, self-efficacy, and career motivation from the pre- to post-questionnaire was different between control group and experimental group after participating with the learning instruction. Grade motivation was no effect of time difference on science motivation in learning.

These findings confirm with previous studies (Tuan et al., 2005) that inquiry lessons can increase students’ learning motivation. In addition (Erhel and Jamet, 2013) found that game can support motivation and learning which offer it with features that immediate learners to actively procedure in the content and (Ebner and Holzinger, 2007) also found that learning will not be successful if there is a lack of motivation. Therefore we needed some tactics to motivate the students to play the game repeatedly.

5. Conclusion

This study reported impacts of educational computer game on students' perceptions and effects of student-associated game-based open inquiry on students' science motivation. The findings revealed successful of improving students' perceptions and science motivation in context of digital game-based learning experience. This implies that the student-associated game-based open inquiry can be effective in fostering chemistry learning of ionization energy. The results from this study could lead us to conclude that the student-associated game-based open inquiry could be an alternative way for promoting chemistry learning in school science.

Acknowledgements

This work was financially supported by the Institute for the Promotion of Teaching Science and Technology (IPST) and Graduate School, Khon Kaen University. The author would like to express gratefully acknowledge to Faculty of Education, Khon Kean University, for supporting this contribution.
References

The Development and Evaluation of the Online Science Fair Inquiry System based on Scaffolding Design

Li-jen Wang, Chien-yu, Cheng, Chiu-ming, Hu & Ying-Tien Wu*
Graduate Institute of Network Learning Technology, National Central University, Taiwan
*ytwu@cl.ncu.edu.tw

Abstract: Science fair is one of the most common open inquiry activities which can facilitate learners to construct their science knowledge and develop science literacy in school. However, there are a great deal of difficulties and challenges in Taiwan’s science fair. For example, novice teachers may neither effectively guide learners to conduct science fair inquiry activities nor facilitate learners to construct related knowledge. To scaffold teachers’ instruction and students’ learning in science fair inquiry effectively, the “Online Science Fair Inquiry System based on scaffolding design” (OSFIS v 2.0) adapted from (OSFIS v 1.0) was developed in this study. After the development of the system, this study also conducted a series of system evaluations on it. To this end, questionnaire survey was conducted. The participants of the system evaluation in this study were 61 elementary school teachers. The participants expressed highly satisfactory perceived usefulness on teachers’ and students’ scaffolding design. Also, they had high intention to use the OSFIS v 2.0. Moreover, this study reveals that the system may facilitate both teachers and students to understand the process of science fair inquiry, solve the limitation of activities times, record the portfolio during inquiry activities, and complement teachers’ professional knowledge. Finally, some suggestions and implications for teachers to conduct open inquiry activities, system design, and future work are also proposed.

Keywords: Science fair; Inquiry; Online Science Fair Inquiry System, Technology-enhanced inquiry tool; Scaffolding

1. Introduction

There is no doubt that inquiry is the core of modern science education. The major educational goal of inquiry-based teaching or inquiry-based instruction is to help learners study science inquiry skills and enhance the understanding of science inquiry (NRC, 2000). In general, there are five stages of an inquiry activity in science classroom; namely questioning, planning, implementing, concluding, and reporting (Lee et al., 2006). According to the openness of inquiry activities, Bell et al. (2005) categorized four different levels of inquiry activities: confirmation, structured inquiry, guided inquiry, and open inquiry. For science educators, K-12 students are expected to be able to conduct open inquiry. In practice, science fair is the most common open inquiry activity in science classrooms. In many countries, science fair is adopted to help student learn science (Bencze & Bowen, 2009). However, the literature revealed that many teachers may lack of professional knowledge, time, recourses, and assistance when conducting science fair instruction (Anderson, 2002). Only few science teachers know how to guide students to conduct science fair projects or inquiry activities effectively (Justi & Gilbert, 2002). In particular, in recent years, lower quality of the science fair projects conducted by elementary school students have been found in Taiwan. Therefore, how to scaffold elementary school teachers’ instruction and students’ learning in conducting science fair projects is crucial for educators.

Recently, various technology-enhanced inquiry tools have been developed to scaffold inquiry activities for different science users, activities, and openness (Chung, 2012). For user type, IQWST (Investigating and Questioning our World through Science and Technology) and WISE (Web-based Inquiry Science Environment; Linn et al., 2003) were developed for scaffolding inquiry activities in junior and senior high school science curriculum while OSFIS v.1.0 (Online
Science Fair Inquiry System v.1.0; Chung; 2012) was specifically developed for elementary science fair. For the inquiry activity type, IQWST was mainly for in-class inquiry, WISE was not only for in-class inquiry but online inquiry and experimental inquiry. And the OSFIS v.1.0 was for science inquiry which was different from IQWST and WISE. And for the degree of the openness, IQWST was structured inquiry, WISE was guided inquiry, and the OSFIS v.1.0 was learner-centered open inquiry. However, technology-enhanced inquiry tool for scaffolding elementary science fair inquiry especially based on scaffolding design is still not yet available. In order to scaffold elementary school science teachers’ science fair instruction and students’ learning in science fair inquiry, this study aimed to develop a new “Online Science Fair Inquiry System” (OSFIS v 2.0) based on scaffolding design. After the development of the OSFIS v 2.0, this study also conducted a series of system evaluations on it.

2. System development

2.1 Participants of system development

The OSFIS v 2.0 developed in this study aimed to provide a platform for elementary teachers who are interested in personal professional development regarding science fair instruction. They can enhance their professional knowledge by using this platform to guide learners to conduct science fair inquiry activities or facilitate learners to construct related knowledge. The development of this system is coordinated by science education and e-learning researcher, in-service science teacher, and system designer. By combining researcher’s professional knowledge and in-service teacher’s practical experience with system designer’s software skills the system design therefore can be more practical for teachers to use.

2.2 System flow chart

In order to help teachers conduct science fair effectively, five inquiry stages suggested by Lee et al. (2006) are included in this system. As shown in Fig. 1, the student flow chart is composed of a series instruction module. Once the students finish each stage of science fair inquiry, they can submit their work to the work reviewing module. If teacher approves the stage, then students can move on to the next one. If not, they need to revise their work according to their teacher’s comments and resubmit their revised work. After students finish the five stages, they have completed a science fair project.

![Figure 1. System flow chart](image-url)
2.3 System framework

The system framework of the OSFIS v 2.0 consists of five main modules and six databases. The six databases store science expert knowledge, student science fair mission, group discussion, interaction, reflection, and user information database. The five modules include user information, interaction and reflection, collaboration, science fair project, and teacher supervision module.

2.4 Modification of Student’s and teacher’s modules

As shown in Fig. 2, the functions derived from OSFIS v 1.0 are presented with a solid line while new functions of OSFIS v 2.0 are presented with a dotted line. In OSFIS v 2.0, “SF Knowledge” (Science Fair Knowledge) is departed from “SF Management” (Science Fair Management) which may assist students acquire more science fair knowledge. “Learning Process” is also added to help students monitor all learning activities that may help students view their own learning process, including “Previous awarded science fair projects”, “Result History”, “Grading History”, “Log book History”, “Discussion History”, and “Reflection History”. More functions which may help students collaborate and interact with others are added, such as “Sharing Data”, “Acquiring Comments”, “Acquiring Replied Hints”, “Teachers’ Hints”, “Account Registration”, and “Add groups”.

![Figure 2. The comparison of student module between OSFIS v 1.0 and v 2.0](image)

For teacher module, as shown in Fig. 3, the functions derived from OSFIS v 1.0 are presented with a solid line while new functions of OSFIS v 2.0 are presented with a dotted line. In OSFIS v 2.0, “SF Knowledge” (Science Fair Knowledge) is departed from “SF Management” (Science Fair Management). Within this module, two more functions, “Read awarded SF projects” and “Add related links” and are added in order to help teachers acquire more science fair knowledge. In “Monitoring & Management module”, group modification is added. Teachers can modify groups’ learning status according to their progress. “Teaching module” is a brand new module which may help teachers review instructional records. In addition, three more functions, “Remind learning progress”, “Project Evaluation”, and “Add friends” are added in OSFIS v 2.0. Teachers have the privilege to remind students with their learning progress and they can also evaluate each project’s phrasal outcome anytime. Teachers can view other teachers’ science fair projects by adding them as good friends.
2.5 Student/Teacher module and scaffolds

The system is designed to help teachers guide students to conduct science fair projects or inquiry activities by adding scaffolds. As a result, it is necessary to elaborate the connection between the system modules and the scaffolds. The relations between student/teacher module and scaffolds are showed as follows: Teacher’s module and scaffolds (Table 1), Student’s module and scaffolds (Table 2):

Table 1: Teacher module and scaffolds

<table>
<thead>
<tr>
<th>Scaffolds</th>
<th>Content</th>
<th>Teaching Challenges</th>
<th>System Modules</th>
</tr>
</thead>
<tbody>
<tr>
<td>Teacher Conceptual scaffolds (TCS)</td>
<td>Offering related Knowledge of Science Fair instruction</td>
<td>Teachers do not understand science fair’s relevant knowledge and structure.</td>
<td>Teaching knowledge module</td>
</tr>
<tr>
<td>Teacher Procedural Scaffolds (TPS)</td>
<td>The procedure of inquiry instruction</td>
<td>Teachers do not know how to conduct science fairs.</td>
<td>Inquiry structure</td>
</tr>
<tr>
<td>Teacher Metacognitive Scaffolds (TMS)</td>
<td>Making plans</td>
<td>Teachers do not know how to make instruction plans</td>
<td>Monitoring &amp; Management module</td>
</tr>
<tr>
<td></td>
<td>Monitoring and adjustment</td>
<td>Teachers do not know how to monitor their instructional process</td>
<td>&amp; Teaching module</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Teachers do not know how to make sure students can finish phrasal learning missions.</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Teachers do not know how to help students modify phrasal missions in order to achieve learning goals</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Teachers cannot monitor students’</td>
<td></td>
</tr>
</tbody>
</table>
learning process

Reflection

Teachers do not know how to Reflect their teaching abilities

Reflection module

Teacher Strategic Scaffolds (TSS)

Teacher-student interaction

Teachers increase the chances to interact with students asynchronously

Interaction module

Peer interaction

Teachers share their progress with other teachers

Table 2: Student module and scaffolds

<table>
<thead>
<tr>
<th>Scaffolds</th>
<th>Content</th>
<th>Teaching Challenges</th>
<th>System Modules</th>
</tr>
</thead>
<tbody>
<tr>
<td>Student Conceptual Scaffolds (SCS)</td>
<td>Offering related Knowledge of Science Fair instruction</td>
<td>Students do not understand science fair’s relevant knowledge and structure.</td>
<td>Teaching knowledge module</td>
</tr>
<tr>
<td>Student Procedural Scaffolds (SPS)</td>
<td>The procedure of inquiry instruction</td>
<td>Students do not know how to Conduct science fairs.</td>
<td>Inquiry structure</td>
</tr>
<tr>
<td>Student Metacognitive Scaffolds (SMS)</td>
<td>Making plans</td>
<td>Students do not know how to make Instruction plans</td>
<td>Monitoring &amp; Management module &amp; Teaching module</td>
</tr>
<tr>
<td></td>
<td>Monitoring and adjustment</td>
<td>Students do not know how to monitor their instructional process</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Students do not know how to finish phrasal learning missions effectively,</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Students do not know how to adjust phrasal missions in order to achieve learning goals</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Students cannot focus on the current learning tasks</td>
<td></td>
</tr>
<tr>
<td>Reflection</td>
<td>Students do not know how to reflect their learning tasks</td>
<td>Reflection module</td>
<td></td>
</tr>
<tr>
<td>Student Strategic Scaffolds (SSS)</td>
<td>Peer interaction</td>
<td>Group leaders need to arrange tasks</td>
<td>Group collaboration module</td>
</tr>
<tr>
<td></td>
<td>Group members need positive interaction</td>
<td>Group members need positive interaction</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Students have difficulties learning tasks</td>
<td>Teacher-student interaction module</td>
<td></td>
</tr>
</tbody>
</table>

3. Methodology (System evaluation)

3.1 Participants

The participants of this study consisted of 61 elementary teachers (20 males and 41 females) whose teaching experience ranged from 1 to 32 years, with an average of 15.81 years. 38 (62.3%) teachers had the experience of instructing science fairs. 21 (34.4%) teachers had no experience of instruction. In general, most participant teacher in this study did not have enough science fair instructional experience which was perfect for this system evaluation.

3.2 Instruments

In this study, the participant teachers’ perceived usefulness and usability of the OSFIS v 2.0 as well as their willingness of using the OSFIS v 2.0 were evaluated. To this end, the 6 Likert-scale questionnaire developed in Yuen & Ma (2008) was adapted and used in this study. The modified instrument consists of three scales: usefulness (5 items), usability (5 items), and willing of use (4 items). All the alpha reliability values of the three scales are greater than 0.8, and the overall alpha reliability value of the instrument is 0.96 (Table 3).
Table 3: Item numbers, reliability, and sample items of the instrument scales

<table>
<thead>
<tr>
<th>Scale</th>
<th>Item</th>
<th>α</th>
<th>Example Item</th>
</tr>
</thead>
<tbody>
<tr>
<td>ITU</td>
<td>5</td>
<td>0.93</td>
<td>I would like to use OSFIS to conduct science fair.</td>
</tr>
<tr>
<td>PU</td>
<td>5</td>
<td>0.96</td>
<td>I find using OSFIS can enhance my teaching.</td>
</tr>
<tr>
<td>PEOU</td>
<td>4</td>
<td>0.94</td>
<td>It is easy for me to master the operation of OSFIS.</td>
</tr>
</tbody>
</table>

ITU, Intention to Use; PU, Perceived Usefulness; PEOU, Perceived Ease of Use.
Over all α = 0.96

3.3 Data collection

There were three phases of data collection. In the first phase (30 minutes), the authors gave directions of five stages of learning tasks, system operations, and student interface, and teacher interface of the system. In the second phase (30 minutes), teachers freely explored the scaffolding design, student interface, and teacher interface. In the last phase (10 minutes), overall system evaluation, usability of learning tasks and scaffolding design, and teachers’ feedback were evaluated by using a questionnaire developed in this study (Fig 4).

![Figure 4. Data collection procedure](image)

4. Major findings and Conclusions

4.1 Major findings

Table 4 shows that the teachers’ average scores on perceived usability of teachers’ scaffolds are between 4.80 to 4.92. The result was higher than the 6 Likert scale average score (i.e., 3.5). It indicates that participants in this study perceived high usability on teachers scaffolds.

Table 4: Teachers’ average scores on perceived usability of teachers’ scaffolds (n=61)

<table>
<thead>
<tr>
<th>Scaffolds</th>
<th>Mean</th>
<th>S.D.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Teacher conceptual scaffolds (TCS)</td>
<td>4.98</td>
<td>0.66</td>
</tr>
<tr>
<td>Teacher procedural scaffolds (TPS)</td>
<td>4.90</td>
<td>0.78</td>
</tr>
<tr>
<td>Teacher metacognitive scaffolds (TMS)</td>
<td>4.91</td>
<td>0.76</td>
</tr>
<tr>
<td>Teacher strategic scaffolds (TSS)</td>
<td>4.90</td>
<td>0.77</td>
</tr>
<tr>
<td>Overall</td>
<td>4.92</td>
<td>0.69</td>
</tr>
</tbody>
</table>
Table 5 shows that the teachers’ average scores on perceived usability of students’ scaffolds are between 4.90 to 4.98. The result was also higher than the 6 Likert scale average score (i.e., 3.5). It indicates that the participants in this study generally held positive attitude toward the system and were willing to use it.

Table 5: Teachers’ average scores on perceived usability of students’ scaffolds (n=61)

<table>
<thead>
<tr>
<th>Scaffolds</th>
<th>Item</th>
<th>Mean</th>
<th>S.D.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Student conceptual scaffolds (SCS)</td>
<td>4</td>
<td>4.83</td>
<td>0.77</td>
</tr>
<tr>
<td>Student procedural scaffolds (SPS)</td>
<td>3</td>
<td>4.92</td>
<td>0.81</td>
</tr>
<tr>
<td>Student metacognitive scaffolds (SMS)</td>
<td>6</td>
<td>4.83</td>
<td>0.76</td>
</tr>
<tr>
<td>Student strategic scaffolds (SSS)</td>
<td>5</td>
<td>4.80</td>
<td>0.76</td>
</tr>
<tr>
<td>Overall</td>
<td>18</td>
<td>4.83</td>
<td>0.71</td>
</tr>
</tbody>
</table>

4.2 Conclusions

To scaffold teachers’ instruction and students’ learning in science fair inquiry, the “Online Science Fair Inquiry System” (OSFIS v 2.0) based on scaffolding design was developed in this study. After the development of the system, this study also conducted a series of system evaluations on it. The participants expressed satisfactory perceived usefulness and ease of use of the system. Also, they had high intention to use the OSFIS v 2.0. Moreover, this study reveals that the OSFIS based on scaffolding design may facilitate both teachers and students to understand the process of science fair inquiry, solve the limitation of activities times, record the portfolio during inquiry activities, and complement teachers’ professional knowledge. Some teachers also gave concrete suggestions of the system design which may be provided as the future revision of the system. Based on the findings in this study, the OSFIS v 2.0 may be improved and applied to different levels of school in the future work.

Acknowledgment

This study was funded by the National Science Council, Taiwan, ROC, under grant contract number NSC 101-2628-S-008-001-MY3, but the opinions expressed in this article do not reflect the position of the National Science Council.

References


Conference Integrating Technology into Computer Science Education. Working group reports from ITiCSE on Innovation and technology in computer science education, United Kingdom.


The difference in Sudoku puzzle-solving ability between undergraduates and postgraduates

Hong-Mei HUa*, Ling-Jin Li, Li-Sha WANGc & Feng-Kuang CHIANGd

a the faculty of education, Beijing Normal University, China
b the faculty of education, Beijing Normal University, China
c the faculty of education, Beijing Normal University, China
d the faculty of education, Beijing Normal University, China
* hongmeihu1990@gmail.com

Abstract: As Sudoku is sweeping around the world, it seems to be increasingly prevalent that Sudoku can contribute to the cultivation of logical thinking ability. In the study, empirical approach was adopted to investigate differences in Sudoku puzzle-solving ability and metacognitive ability of mathematical problem-solving between undergraduates and postgraduates, and to examine the relationship between Sudoku puzzle-solving ability and metacognitive ability of mathematical problem-solving. All participants in the study were students studying in Beijing Normal University. The results indicate: (a) no difference between undergraduates and postgraduates in solving the same level Sudoku puzzles and metacognitive ability of mathematical problem-solving, and (b) metacognitive ability of mathematical problem-solving having no significant effect on Sudoku puzzle-solving ability. However, the number of Sudoku puzzles participants had ever finished had appreciable impact on Sudoku puzzle-solving.

Keywords: Sudoku, undergraduates, postgraduates, metacognitive ability

1. Introduction

Sudoku is a logic-based combinatorial (Lenstra, Kan, & Shmoys, 1985) number-placement puzzle. The objective is to fill a 9×9 grid with digits so that each column, each row, and each of the nine 3×3 sub-grids that compose the grid contain all of the digits from 1 to 9. The puzzle setter provides a partially completed grid, and every puzzle has a unique solution (Wayne, 2006).

Problem-solving is a kind of mental activity of seeking methods to deal with problems when facing new situations and new tasks or realizing a lack of ready-made countermeasure to deal with the contradiction between subjective and objective needs in daily life and social practice. Mathematical problem-solving is one of the most important basic processes in mathematical thinking. Researchers defined mathematical problem-solving from the perspective of cognitive theories that it’s a process of a series of perceiving (information processing) which depend on cognitive operator with the guidance of target problem and the information provided by the problem (Jinqiu, 1995).

As to metacognition, different scholars define it differently. Metacognition theory considers human learning a procedure of not only conceiving, memorizing, understanding and processing memory materials but self-perceiving, controlling and modulating cognition process. All in all, metacognition is participants perceiving their cognitive activities (Xueying, 2008). Berardi-coletta et al. (1995) tested college students by a traditional question named tower of Hanoi and got a conclusion that metacognition training is good for solving problems. Ping Yu’s (2002) research indicated that metacognition level shows a close correlation with mathematic problem-solving ability for senior high school students. Besides, there are researches which indicate that we can improve students’ mathematical problem-solving ability by developing their metacognitive ability (Yan, 2005).

Sudoku puzzles are popular all around the world. According to the media that children can improve their logical thinking ability and develop intelligence by solving Sudoku puzzles, while middle-aged people can bring back their energetic mind and throw away life pressure (Song, 2005). Currently some schools of China organize students to solve Sudoku puzzle at math class regularly, expecting to improve students’ math scores. But can Sudoku puzzles really improve students’ math scores? In the existing literature there are a lot about strategies and arithmetic in solving Sudoku puzzles such as a 9×9 solving strategy introduced by Davis (2006). Some studied how Sudoku puzzles influence
humans. For example, Baek, et al. (2008) indicated that Sudoku puzzles can help people develop their logical reasoning ability. There are other researchers focusing on other topics with Sudoku puzzles. Such as Chuen-Tsai Sun et al. (2011), they used Sudoku puzzles as the digital scaffolding to study problem-solving behavior with the premise that if a player can solve the Sudoku puzzles from one level to a higher one, his ability is improved in solving Sudoku puzzles.

However, there is more experience than empirical studies about weather Sudoku puzzles can lead to those changes. No information illustrates a specific function of Sudoku puzzles neither facilitation in students’ math scores and mathematical problem-solving ability.

This study aims at probing the relationship between metacognitive ability of mathematical problem-solving and Sudoku puzzle-solving ability. By comparing science undergraduates and postgraduates’ metacognitive ability of mathematical problem-solving and Sudoku puzzle-solving ability, we expect to reveal the relationship between the two kinds of ability and explore whether Sudoku puzzles has a close correlation with math scores and mathematical problem-solving ability. The follows are hypotheses of the study.

(a) There is significant difference in metacognitive ability of mathematical problem-solving between undergraduates and postgraduates.
(b) There is significant difference in Sudoku puzzle-solving ability between undergraduates and postgraduates.
(c) Metacognitive ability of mathematical problem-solving has appreciable effect on Sudoku puzzle-solving.
(d) Average math scores in senior high school have positive correlation with Sudoku puzzle-solving ability.

2. Methods

2.1 Participants

30 students of Beijing Normal University were selected by convenient sampling, 15 science undergraduates and 15 science postgraduates. 29 participants came from the school of educational technology, 1 from mathematics. As to gender, postgraduate group consists of 5 males and 10 females, and undergraduate group consists of 4 males and 11 females.

2.2 Materials

2.2.1 Metacognitive ability questionnaire

Two parts form the questionnaire, one part for fundamental information including participants’ gender, major, the number of done Sudoku puzzles before the study and average score of math in high school, and the other part for testing the participants’ metacognitive ability of mathematical problem-solving which is a scale including 37 items.

The 37 items come from a metacognition scale in mathematical problem-solving compiled by Jianlan Tang et al. of Guangxi Normal University in 2005, which is a comprehensive analysis of Panaaela’s questionnaire, Skalower and Shbling’s questionnaire, Jianyou Zhang’s self-monitoring ability questionnaire in math among middle school students, Ping Yu’s self-monitoring ability questionnaire in mathematical problem-solving, and Haiyan Guo’s dynamic and static metacognitive questionnaire with a series of unstructured questionnaire verbal reports in problem solving process. This scale is a structured original questionnaire adopting a five Likert scale. In the scale, Metacognitive ability is divided into 3 main factors and 37 items. The main factors are metacognitive knowledge, metacognitive experience, and metacognitive strategy. The sum score of these 37 items a participant get is seemed as metacognitive ability score. The scale’s α coefficient of the total internal consistency is 0.901, indicating a good reliability and validity (Jianlan, Ying & Fucheng, 2005).

As mentioned earlier, metacognitive ability has a close correlation with mathematic problem-solving ability (Ping, 2002. Yan, 2005). In the study, we regarded the metacognitive ability score as participants’ mathematic problem-solving ability.
2.2.2 Sudoku puzzles

Sudoku is a number-placement puzzle with the rule that each column, each row, and each of the nine 3x3 sub-grids that compose the grid contain all of the digits from 1 to 9 but no repetition is allowed.

We selected 9 basic Sudoku puzzles from Sudoku puzzles (2006) compiled by Wayne Gould in New Zealand. Every puzzle is presented on a half of A4 paper.

The maximum score of each puzzle is 10. The score of each puzzle a participant got is according to the participant finished percentage in total blanks. A participant’s Sudoku score is the sum of 9 Sudoku puzzles’ scores he got in 1 hour. In the study, we regarded the Sudoku score a participants got as his/her Sudoku puzzle-solving ability.

2.2.3 Interview question

A brief interview question was designed to investigate participants’ attitude towards Sudoku puzzles. The question is “Do you want to do Sudoku puzzles in the future?”

2.3 Research process

In order to make the study more reasonable and achievable, a pilot study was conducted in which 2 undergraduates participated, and then we revised some details of the study with their advice. The 15 undergraduates didn’t include the 2 undergraduates participated in pilot study. Every participant decided specific time respectively from Dec. 20th, 2013 to Jan. 6th, 2014. It took each participant about 70 minutes. Below was the detailed process.

(a) A researcher explained to the participant the study content, purpose and what they would do.
(b) The participant filled in the metacognitive ability questionnaire. It took about 3 minutes.
(c) A researcher gave the 9 Sudoku puzzles to the participant, explained what was Sudoku and the rule, and remind the participant that he/she could choose a few to do, not all.
(d) The participant got down to Sudoku puzzle-solving for 1 hour with pencils and erasers.
(e) A researcher asked the participant the interview question and recorded the answer.

3. Results

SPSS was used to analyze the data of the study. Significance level was 0.05.

3.1 Differences in metacognitive performance and Sudoku puzzle-solving performance

Independent-Samples T test was used to test the differences in metacognitive performance and Sudoku puzzle-solving performance between undergraduates and postgraduates (Table 1, Table2).

<table>
<thead>
<tr>
<th>Dimension</th>
<th>Group</th>
<th>N</th>
<th>Mean</th>
<th>SD</th>
<th>Sig.(two-tailed)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Metacognitive knowledge</td>
<td>Undergraduates</td>
<td>15</td>
<td>27.00</td>
<td>6.36</td>
<td>0.665</td>
</tr>
<tr>
<td></td>
<td>Postgraduates</td>
<td>15</td>
<td>26.13</td>
<td>4.31</td>
<td></td>
</tr>
<tr>
<td>Metacognitive experience</td>
<td>Undergraduates</td>
<td>15</td>
<td>19.40</td>
<td>5.53</td>
<td>0.944</td>
</tr>
<tr>
<td></td>
<td>Postgraduates</td>
<td>15</td>
<td>19.53</td>
<td>4.75</td>
<td></td>
</tr>
<tr>
<td>Metacognitive strategy</td>
<td>Undergraduates</td>
<td>15</td>
<td>60.27</td>
<td>12.31</td>
<td>0.447</td>
</tr>
<tr>
<td></td>
<td>Postgraduates</td>
<td>15</td>
<td>63.40</td>
<td>9.78</td>
<td></td>
</tr>
<tr>
<td>Total score</td>
<td>Undergraduates</td>
<td>15</td>
<td>106.67</td>
<td>21.52</td>
<td>0.729</td>
</tr>
<tr>
<td></td>
<td>Postgraduates</td>
<td>15</td>
<td>109.07</td>
<td>15.64</td>
<td></td>
</tr>
</tbody>
</table>

As shown in Table 1, the difference of undergraduate and postgraduates’ total mean scores is 2.40. The mean scores of undergraduate and postgraduates are very close in metacognitive knowledge.
and metacognitive experience dimensions, and score differences in metacognitive strategy dimension is 3.13. The standard deviations of postgraduates’ score in metacognitive knowledge, metacognitive experience and metacognitive strategy are less than those of undergraduates. Overall, the results indicate that no statistical significance (p>0.05) in metacognitive ability between undergraduates and postgraduates, and invalidate hypothesis a.

Table 2: Two-tailed t-test results for undergraduates and postgraduates’ Sudoku scores.

<table>
<thead>
<tr>
<th>Group</th>
<th>N</th>
<th>Mean</th>
<th>SD</th>
<th>Sig.(two-tailed)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Undergraduates</td>
<td>15</td>
<td>24.86</td>
<td>11.91</td>
<td>0.656</td>
</tr>
<tr>
<td>Postgraduates</td>
<td>15</td>
<td>27.14</td>
<td>15.57</td>
<td></td>
</tr>
</tbody>
</table>

As shown in Table 2, the difference of undergraduates and postgraduates’ Sudoku mean scores is 2.28. The results indicate no statistical significance (p>0.05) in Sudoku puzzle-solving ability between undergraduates and postgraduates, so hypothesis b is invalid.

3.2 Correlations of Sudoku score with other variables

The linear regression analysis was used to analyze the correlations of Sudoku score with other variables (Table 3).

Table 3: Coefficienta of linear regression results for correlations of Sudoku scores with other variables.

<table>
<thead>
<tr>
<th>Model</th>
<th>Unstandardized Coefficients</th>
<th>t</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>B</td>
<td>Standard Error</td>
<td></td>
</tr>
<tr>
<td>(Constant)</td>
<td>-8.721</td>
<td>25.537</td>
<td>-0.342</td>
</tr>
<tr>
<td>Sex</td>
<td>3.448</td>
<td>5.322</td>
<td>0.648</td>
</tr>
<tr>
<td>Grade</td>
<td>2.482</td>
<td>5.012</td>
<td>0.495</td>
</tr>
<tr>
<td>The number of done Sudoku puzzles</td>
<td>8.689</td>
<td>3.108</td>
<td>2.796</td>
</tr>
<tr>
<td>Average math scores in senior high school</td>
<td>1.377</td>
<td>4.865</td>
<td>0.283</td>
</tr>
<tr>
<td>Metacognitive knowledge</td>
<td>1.701</td>
<td>1.093</td>
<td>1.556</td>
</tr>
<tr>
<td>Metacognitive experience</td>
<td>-1.710</td>
<td>1.141</td>
<td>-1.499</td>
</tr>
<tr>
<td>Metacognitive strategy</td>
<td>-0.041</td>
<td>0.286</td>
<td>-0.143</td>
</tr>
</tbody>
</table>

Data in Table 3 indicates that only the number of done Sudoku puzzles before the study has a significant effect on Sudoku score (p<0.05), and metacognitive ability has no significant effect on Sudoku puzzle-solving ability. Hypothesis c is invalid.

The linear regression analysis was used to analyze the correlations of Sudoku score with the number of done Sudoku puzzles before the study, as ANOVA results shown in Table 4.

Table 4: ANOVAa results for correlations of Sudoku scores with the number of done Sudoku puzzles.

<table>
<thead>
<tr>
<th>Model</th>
<th>Sum of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regression</td>
<td>1379.710</td>
<td>1</td>
<td>1379.710</td>
<td>9.567</td>
<td>0.004b</td>
</tr>
<tr>
<td>Residual</td>
<td>4030.049</td>
<td>28</td>
<td>144.216</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>5417.759</td>
<td>29</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Data in Table 4 shows that the number of done Sudoku puzzles before the study can be used to predict Sudoku score. The linear regression equation is \( Y = 14.246 + 7.668X \).

Wenling Li et al. (2008) divided correlativity into four levels according to Pearson correlation coefficient: (a) No correlation or weak correlation, (b) normal correlation, (c) high correlation, and (d) strong correlation. Pearson correlation coefficient for all variables shows that Pearson correlation coefficient of Sudoku score and the number of done Sudoku puzzles before the study is 0.505, which
means the two variables have high correlativity. Pearson correlation coefficient of average math scores in senior high school and Sudoku score is 0.092, which means that these two variables have no correlativity. So hypothesis d is invalid.

3.3 Interview results

According to the results of metacognitive ability questionnaire, there were 20 participants hadn’t done any Sudoku puzzles, 11 undergraduates, 9 postgraduates. 17 participants who hadn’t done any Sudoku puzzles before answered this question. 12 of them said they wanted to do, 3 said they didn’t want to do any more, 2 said they might want to do. 9 participants who had done some Sudoku puzzles before answered this question. 5 of them said they wanted to do.

4. Discussion

4.1 Average math scores in senior high school and Sudoku score

On the basis of this study results, there is no correlation between average math scores in senior high school and Sudoku score. So that we doubt whether it is worth students’ while solving Sudoku puzzle at math class regularly. Class time is very important for students. If solving Sudoku puzzle cannot improve students’ ability or scores, maybe school should reconsider how to help students make better use of class time.

However, we cannot draw the conclusion that solving Sudoku puzzles can’t improve math abilities as this study has several limits. (a) This study only include average math scores in senior high school, not including math scores in junior high school and primary school. So it is illogical to say that solving Sudoku puzzles cannot improve math scores. (b) We divided average math scores in senior high school into five score section which is not a subtle rational division. (c) We gave only 9 puzzles to participants and they did Sudoku puzzles just for 1 hour. They cannot form a steady Sudoku puzzle-solving ability in such a short time, so the data maybe represents participants’ ability accurately.

In the future, we can conduct a long-term tracking study in primary school and junior high school to explore whether solving Sudoku puzzles has impact on math scores.

4.2 The number of done Sudoku puzzles before the study and Sudoku score

According to the result, the number of done Sudoku puzzles before the study has a significant impact on Sudoku score. We can infer that to some extent, the larger number of Sudoku puzzles a participant had done the more experience and strategies he might get, and he would get the higher Sudoku score in the study. However, we should consider whether a certain number may exist. Participants may get equivalent scores when they had done more than the certain number of Sudoku puzzles. This requires a deeper study.

4.3 Metacognitive ability and Sudoku score

The results show that there is no significant difference in metacognitive ability between undergraduates and postgraduates. A possible reason is that two groups of participants are adults and in the stage of formal operational stage defined by Piaget. They have a steady cognitive level, and there is no significant difference in metacognitive ability among them. Another possible reason is that the metacognitive ability questionnaire doesn’t work well to detect the metacognitive level.

Data show that metacognitive ability has no statistical significant effect on Sudoku score. However, we can’t draw the conclusion that metacognitive ability has no significant impact on Sudoku puzzle-solving ability, because we don’t have clear evidence that metacognitive ability of mathematical problem-solving is the same with metacognitive ability in Sudoku puzzle-solving. The metacognitive ability questionnaire detects the former not the latter. It is necessary to design a special scale to test metacognitive ability in Sudoku puzzle-solving in further study.
4.4 Attitude towards Sudoku puzzles

According to results of interview question, although 71.4 % of participants who had never done Sudoku puzzles before the study, 70.6% of them wanted to try it later. So we can speculate that Sudoku puzzle is an attractive game. Why is it so attractive? And why do some people think solving Sudoku puzzles is good for them improving ability subjectively? We can do another study to investigate the reason.

5. Conclusion

According to the study, there is no significant difference in metacognitive ability of mathematical problem-solving between undergraduates and postgraduates, and no significant difference in Sudoku puzzle-solving ability between undergraduates and postgraduates. Besides, metacognitive ability of mathematical problem-solving has no appreciable effect on Sudoku puzzle-solving. In addition, average math scores in senior high school have no positive correlation with Sudoku puzzle-solving ability. However, the number of done Sudoku puzzles before the study has a significant effect on Sudoku puzzle-solving ability.

There were limitations in the study. We had a small number of participants and more proper materials should be used.

Acknowledgements

This study thanks sincerely to professor Keding Zhong for his guidance and suggestion to study goals and data analysis method, Yehui Wang for detailed advice on data analysis and discussion, and vice professor Yin Wang for her kindness to provide research facility. Besides, we thank to the 32 participants.

References


Investigating the role of self-explanation and co-explanation in 4th graders’ game-based science learning

Chung-Yuan HSU\textsuperscript{a*}, Hung-Yuan WANG\textsuperscript{b}, & Shih-Hsuan WEI\textsuperscript{c}

\textsuperscript{a}Department of Child Care, National Pingtung University of Science and Technology, Taiwan
\textsuperscript{b}Graduate Institute of Applied Science and Technology, National Taiwan University of Science and Technology, Taiwan
\textsuperscript{c}Department of Mathematics Education, National Taichung University of Education, Taiwan

*jackohio@gmail.com

Abstract: This study extends our previous studies on investigating the effects of embedding self-explanation principle into game-based science learning. In order to enhance the students’ generating their own explanations during the game, we replaced the multiple choice questions with the design of allowing dyads to co-explain their causes of failure in the game via utilizing the technique of online chat. The participants were 60 4th graders recruited from an elementary school in southern Taiwan. They were randomly assigned to dyads of either an experimental group (conducting co-explanation via online chat) or a control group (conducting self-explanation via multiple choice questions). The measurements included the pretest, posttest, and a three-week retention test. The results show that both games had a positive impact on facilitating the students’ acquisition of scientific concepts. But, the players who performed co-explanation via online chat did not outperform those who used multiple choice questions as self-explanation prompts. Through analysis of dialogue of the players in the experiment group, we found that the quality of the dyads’ dialogue was poor; they rarely discussed the causes of failure when the prompts appeared in the game.

Keywords: Game-based learning, self-explanation, science learning, multiplayer game

1. Introduction

Researchers (Chi, Bassok, Lewis, Reimann, & Glaser, 1989) found that students learned well when they were asked to generate explanation to themselves. This constructive learning process enables learners to generate inferences to fill in information gap, integrating information, and monitoring and repairing faulty knowledge (Roy & Chi, 2005). In the recent years, a growing number of researchers attempt to integrate self-explanation principle into educational games and investigate its impacts on players’ learning outcomes (Adams & Clark, 2014; Hsu, Tsai, & Wang, 2012 & Hsu, Tsai, & Wang, in press; Johnson & Mayer, 2010). This study extends our previous studies (Hsu et al., 2012, in press) on investigating the effects of embedding self-explanation principle into game-based science learning. Although both studies as well as the previous research (Adams & Clark, 2014; Johnson & Mayer, 2010; O’Neil et al., 2014) have identified the positive impacts of using multiple choice questions as self-explanation prompts, it might still risk limiting learners’ generating inferences and hinder robust learning outcomes. Thus, to enhance the students’ generating their own explanations, Hsu et al. (in press) suggested replacing the multiple choice questions with the design of allowing dyads to co-explain their causes of failure in the game via utilizing the technique of online chat. Through interaction with peers in the game, we hypothesize that the experimental condition would outperform the control condition since the dyads in the experimental group could share diverse perspectives, co-construct knowledge, and benefit from explaining another person’s reasoning. In sum, this study attempted to examine how different forms of self-explanation influence students’ game-based science learning. The guiding questions are:

1. What are the effects of self-explanation and co-explanation in game-based science learning?
2. How is the quality of co-explanation during game playing?
2. Methodology

2.1 Participants

The participants were 60 4th graders recruited from an elementary school in southern Taiwan. Without receiving formal instruction regarding light and shadow concepts, they were randomly assigned to dyads of either an experimental group (conducting co-explanation via online chat) or a control group (conducting self-explanation via multiple choice questions). There were 13 females and 17 males in the experimental condition and 16 females and 14 males in the control condition. Both groups played a multiplayer game with self-explanation embedded.

2.2 The game

The game of this study was developed by the researchers to support forth graders’ learning of shadow and light concepts. The game consisted of three stages and each one was designed to instruct a core concept, such as the relationship between the height of a light source and the length of the shadow produced, shadow change throughout the day, and shadow intensity, respectively for Stage 1 to 3. The participants were required to play the game with a peer randomly assigned by the researchers (see Figure 1). That is, neither of them knew who their partner was or where she or he was situated. During game playing, a self-explanation prompt appears whenever a mistake is made. Both players had to stop playing and respond to the prompt. The participants in the experimental group were encouraged to discuss the causes of failure via online chat. When the discussion was completed, they could click a button and continue the game. However, the students in the control group used multiple-choice questions as self-explanation prompts in the game context. The time limitation for all the three stages was 35 minutes. The players would be directed to the posttest when failing to meet the limitation.

![Figure 1. Screenshot of the game.](image)

2.3 Measurement

A 10-item test was used to measure the participants’ understanding regarding light and shadow covered around the main concepts in the game. Each student took the test before the game, right after the treatment, and three weeks after the treatment. Sharing the same questions, the test only varied in the order of displaying the questions and options. These items were also used in Hsu et al.’s (2011, in press) studies. The reliability coefficient was 0.60 in Hsu et al. (in press), suggesting acceptable reliability.
2.4 Procedure

Participants were randomly assigned to either the experimental or control group and individually seated at a computer when entering the computer classroom. A researcher introduced the study and the tasks to the class. Following the introduction, the students took a pretest without a time limit (averagely less than six minutes). Later, the researchers helped the students build up an online connection of the game with their partners, and log in Skype (a technology allows users to communicate with peers by using a microphone over the Internet) for those in the experimental condition. Their narration during the game playing would be recorded for further analysis. The students then played the game for 30 minutes. They received a posttest when passing the three stages or over the time limit. Each student also took a retention test after three weeks.

2.5 Data analysis

First of all, a series of paired t-test were conducted to compare students’ improvement from the pretest. This study later investigated the score difference of both groups by using the pretest scores as a covariate. A content analysis was utilized to probe the dialogue of the players in the experimental group during game playing.

3. Results

Table 1 shows the results of paired t-tests. As indicated, the students’ posttest and retention scores were significantly higher than their pretest scores in both experimental and control condition. This finding suggests that both games could positively facilitate the students’ acquisition of scientific concepts.

<table>
<thead>
<tr>
<th>Type</th>
<th>Group</th>
<th>N</th>
<th>Adjusted mean</th>
<th>Std. error</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Posttest</td>
<td>Control</td>
<td>30</td>
<td>7.61</td>
<td>.29</td>
<td>2.57</td>
</tr>
<tr>
<td></td>
<td>Experimental</td>
<td>30</td>
<td>8.26</td>
<td>.29</td>
<td></td>
</tr>
<tr>
<td>Retention</td>
<td>Control</td>
<td>30</td>
<td>7.47</td>
<td>.33</td>
<td>1.44</td>
</tr>
<tr>
<td></td>
<td>Experimental</td>
<td>30</td>
<td>8.03</td>
<td>.33</td>
<td></td>
</tr>
</tbody>
</table>

This study further examined the score difference between the two groups by using the pretest scores as a covariate and the posttest and retention score as dependent variables. The assumption of homogeneity of regression was tested and was not violated ($F=1.16, p>.05; F=1.44, p<.05$). The ANCOVA results of the posttest and retention are shown in Table 2. As shown, no statistically significant difference was identified. That is, the players who performed co-explanation via online chat did not outperform those who used multiple choice questions as self-explanation prompts.

Table 2: Descriptive data and ANCOVA results for the posttest and retention scores.

As aforementioned, the players’ communication during the game would be recorded and transcribed for further analysis. In this preliminary analysis, we focus on the players’ narration right
after the failure in the game. The results show that the dyads rarely discussed the causes of failure when
the self-explanation prompt appeared, such as:
Participant 12: I am dead.
Participant 10: I am dead, too.
Participant 12: No problem, let’s play again.
Participant 10: Well, this time we should walk slowly.

In addition, they tended to blame their partner for the cause of mistakes. Take Participant 13 for
instance, “I hate you. I only make one mistake but you make two. It is annoying that we keep failing.”
Although some dyads might come up with the tricks to pass the game, these tricks were not absolutely
correct. An example is:
Participant 1: Oops, I am completely dead
Participant 6: I told you not to move but you never listen. Maybe you should walk on the red
lane. Be careful! Do not fall in the sea.

Regarding the above example, the players should pay attention to shadow change throughout
the day, rather than the difference in the lanes they walk.

4. Discussion

Self-explanation effects become effectively when learners can generate inferences to fill in missing
information, integrate information and repair faulty knowledge (Roy & Chi, 2005). However, the
previous research pointed out that utilizing multiple choice questions as self-explanation prompts in the
game context was likely to limit the players’ generating inferences (Hsu et al., 2014). To solve this
problem, this study implemented a design by having dyads co-explain their causes of failure during
game playing and investigate its impact on the participants’ learning outcomes. However, no
statistically significant difference was identified. Players who co-explained via online chat did not
perform better than those who used multiple choice questions as self-explanation prompts. In addition,
through analyzing the dyads’ dialogue, we found that the quality of the dyads’ narration was not
satisfied and they rarely discussed the possible causes of failure when the prompt appeared. They
chaanted all the time and blamed their partner for failure. Although some of them could identify some
tricks to pass the game, they might not be accurately linked to the targeted concept.

According to Chi’s (2009) framework of passive-active-constructive-interactive learning
strategies, interacting with a peer in a computer-based environment can be classified as interactive
learning activities only when the dialogue includes substantive contributions from both partners, as well
as learners respond to scaffoldings and modify errors based on feedback. It seems like that the
participants of the experimental condition simply taking turns speaking, which could not be categorized
as an interactive learning event. To sum up, having dyads collaboratively construct knowledge in
game-based science learning is a one of ultimate level of learning strategies. But, future studies still
need to think about ways to promote quality of players’ interaction, such as designing events to confront
or challenge the partner’s statements, or encourage involvement into deeper discussion.

Acknowledgements

This research was supported by the projects from the National Science Council, Republic of China,
under contract number NSC 102-2511-S-020-001 and MOST 103-2511-S-020 -001 -MY3.

References

environment: Keeping gaming in motion. Computers & Education, 73, 149-159. doi:
10.1016/j.compedu.2014.01.002


Preliminary Requirements Analysis towards an Integrated Learning Analytics System

Byung-gi CHOIa, Yong-sang CHOOb, Jaeho LEEc*
aSchool of Electrical and Computer Engineering, University of Seoul, Republic of Korea
bKorea Education & Research Information Service
cSchool of Electrical and Computer Engineering, University of Seoul, Republic of Korea
*jaeho@uos.ac.kr

Abstract: An integration of various information and processes for learning analytics into a united framework is the key to the development of an open and extensible learning analytics system. Recently, we have taken a step towards developing such a framework by starting to build a reference software architecture which in turn will allow us to identify the structure and the workflow of learning analytics systems. Our final goal is to develop an explicit specification of the learning analytics architecture as the international standard so that open and extensible learning analytics systems can be built for worldwide interoperability. In this paper, we present the result of our preliminary requirement analysis towards such an open and interoperable learning analytics system. The analysis focuses mainly on the system aspects of existing well-known frameworks such as IMS Global learning analytics platform.

Keywords: Learning Analytics, Requirement Analysis, Reference Architecture, Standards

1.Overview

Learning analytics (LA) systems require integration of the processes of measurement, collection, analysis and reporting of data about learners and their contexts, and thus involve multi-disciplinary areas including artificial intelligence, information science, statistics, visualization, and so on. An effective integration of various information and processes for learning analytics into a united framework is the key to the development of an open and extensible learning analytics system because the system should be open to several related areas of research such as academic analytics, action research, educational data mining, personalized adaptive learning, and more. The system also should be extensible as new methodologies and technologies emerge rapidly. Especially, the big data technology has been evolved markedly to make it possible to collect data massively, analysis instantly, and visualize appropriately for learning analytics field.

Recently, we have taken a step towards developing such a framework by starting to build a reference software architecture which in turn will allow us to identify the structure and the workflow of learning analytics systems.

Our final goal is to develop an explicit specification of the learning analytics architecture as the international standard so that open and extensible learning analytics systems can be built for worldwide interoperability. In this paper, we present the result of our preliminary requirements analysis towards such an open and interoperable learning analytics system. The analysis focuses mainly on the system aspects of existing well-known frameworks such as IMS Global learning analytics platform.

The remainder of this paper is organized as follows. Firstly, we briefly review the purpose of learning analytics and survey related works and standardization activities in learning analytics field. Based on this survey, we then discuss the basic requirements of the reference software architecture. Finally we gives a summary of the main results of this paper and highlight directions for future work.
2. Backgrounds

In this section, we introduce the basic backgrounds that are needed for requirement analysis for learning analytics including the analysis levels of learning analytics and standardization activities.

2.1 Learning Analytics and Big Data

Learning analytics, as stated in the “Call for Papers of the 1st International Conference on Learning Analytics & Knowledge (LAK 2011)”, is “the measurement, collection, analysis and reporting of data about learners and their contexts, for purposes of understanding and optimizing learning and the environments in which it occurs.” The tasks of measurement, collection, analysis, and reporting in this definition correspond closely to the major activities in big data, that is, collection, processing, analysis, and visualization of data, as shown in Figure 1.

![Figure 1. Big data workflow](image)

Such correspondence is not coincidental but suggests that learning analytics can take advantage of the technological advancement of big data in building a learning analytics framework. Our requirement analysis for learning analytics is also largely borrowed from that of the big data framework.

2.2 Range of Learning Analytics

According to Buckingham Shum (2012), the range of learning analytics can be defined as macro-, meso- and micro-levels.

![Figure 2. Levels of Learning Analytics (Buckingham Shum, 2012)](image)
Buckingham Shum (2012) describes the levels of learning analytics as follows.

- **Macro-level analytics** is the cross-institutional analytics over region, state, national, or international institutions for students’ lifetimes. Macro-analytics becomes increasingly real-time, incorporating more data from lower meso- or micro-levels, utilizing data integration methodologies that are developed in non-educational sections.

- **Meso-level analytics** operates at institution level, benefiting from the common business processes for business intelligence, utilizing the tools to integrate data silos in enterprise warehouse, optimize workflows, generate dashboards, mine unstructured data, predict future trends, and so forth.

- **Micro-level analytics** supports the tracking and interpretation of process-level data for individual learners or groups. This data is of primary interest to learners themselves and correspondingly the most personal, since it can disclose online activities as well as physical activities such as geolocation, library loans, purchases, and interpersonal data such as social networks.

As shown in Figure 2, while the aggregation of user data from the micro-levels enriches meso- or macro-level analytics, the breadth and depth at the macro- or meso-levels add power to micro-level analytics in building predictive models or providing feedback to learners. We thus believe that an effective integration of data and activities among these layers is essential requirement for mutual enrichment.

### 2.3 Reference Model for Learning Analytics

Chatti, Dyckhoff, Schroeder, and Thüs (2012) describe a reference model for learning analytics based on four dimensions: data and environments (what?), stakeholders (who?), objectives (why?), and methods (how?) as depicted in Figure 3.

![Learning Analytics Reference Model](image)

This reference model is going to be utilized in our requirement analysis as it provides a classification schema on software components in our ongoing development of reference software architecture.
2.4 Standardization Activities

As the importance of the interoperability of learning analytics is recognized in Korea, a three-year-long research project funded by Telecommunications Technology Association (TTA) of Korea is recently launched to build a reference model of learning analytics based on educational content and unstructured data. In this project, Korea Education and Research Information Service (KERIS) runs the project in close coordination with ISO/IEC JTC 1/SC 36 and IMS Global.

ISO/IEC JTC 1/SC 36, Information Technology for Learning, Education and Training is a standardization subcommittee (SC), which is part of the Joint Technical Committee ISO/IEC JTC 1, that develops and facilitates standards within the field of information technology (IT) for learning, education and training (LET) to support individuals, groups, or organizations, and to enable interoperability and reusability of resources and tools. Recently, SC 36 established an Ad Hoc Group on Learning Analytics Interoperability to ascertain the necessity of standards to facilitate interoperability among diverse learning analytics components such as data collection, analysis, privacy protection, qualification of data and accessibility. We expect to submit the result of requirement analysis to this Ad Hoc Group as contributions for assessment, and anticipate to participate in the future Working Group, if established, with the reference software architecture.

3. Related Work

3.1 IMS Global

IMS Global Learning Consortium (usually referred to as IMS GLC, IMS Global or simply IMS) is a global, nonprofit, member organization that strives to enable the growth and impact of learning technology in the education and corporate learning sectors worldwide. IMS GLC members provide leadership in shaping and growing the learning industry through community development of interoperability and adoption practice standards and recognition of the return on investment from learning and educational technology. Their main activity is to develop interoperability standards and adoption practice standards for distributed learning, some of which like Learning Tools Interoperability (LTI), Question & Test Interoperability/Accessible Portable Item Protocol (QTI/APIP), Common Cartridge, Learning Information Services and Content Packaging are very widely used (Wikipedia, 2014a).

Especially, IMS Caliper (IMS Global Learning Consortium, 2013) is a work in progress to define a learning measurement framework using existing IMS specifications to provide a standardized representation, capture, and marshaling of learning activity generated metrics targeted for consumption by any conforming sensor API endpoint enabled analytics store/services. IMS Caliper is built around the following concepts:

- **IMS Learning Metric Profiles** that provide a Learning Activity centric focus to standardize on metrics (actions and related context) captured across consumer and producer learning tool’s delivery activities and delivery platforms that consume and orchestrate activity based curriculum, while providing for custom extensions and future additions to the profiles;
- **IMS Learning Sensor API and Learning Events** drive standardized instrumentation and metric capture and marshal between tools and their delivery platforms and/or associated analytics service solution aggregating metrics;
- **IMS LTI™/LIS/QTI™ leverage and extensions** enhance and integrate granular, standardized learning measurement with tools interoperability and the underlying learning information models, inclusive of course, learner, outcomes and other critical associated context.

In a recent IMS publication, “Learning Measurement for Analytics Whitepaper” (IMS Global Learning Consortium, 2013), they claim standards for learning analytics are required so they can be combined across all of the educational sources by asserting that “equipped with a standards based common foundation for learning measurement, the quality, efficacy and performance derived analytics for the online curriculum across the ecosystem can be achieved more effectively.”
UNESCO Institute for Information Technologies in Education (IITE) identifies three kinds of predictors and indicators, and two kinds of interventions as follows.

- **LMS/VLE Analytics Dashboards**: The first kinds of analytics that many institutions will encounter will be the analytics dashboards now appearing in most online learning platforms. Data logs are now rendered via a range of graphs, tables and other visualizations, and custom reports designed for consumption by learners, educators, administrators and data analysts.

- **Predictive Analytics**: From the pattern of learners’ static data (e.g. demographics; past attainment) and dynamic data (e.g. pattern of online logins; quantity of discussion posts) one can classify the trajectory that they are on (e.g. “at risk”; “high achiever”; “social learner”), and hence make more timely interventions (e.g. offer extra social and academic support; present more challenging tasks).

- **Adaptive Learning Analytics**: Adaptive learning platforms build a model of a learner’s understanding of a specific topic (e.g. algebra; photosynthesis; dental surgical procedures), sometimes in the context of standardized tests which dictate the curriculum and modes of testing. This enables fine-grained feedback (e.g. which concepts you have grasped and at what level), and adaptive presentation of content (e.g. not showing material that depends on having mastered concepts the learner has failed on).

- **Social Network Analytics**: Social network analysis (sometimes called Organizational Network Analysis in corporate settings) makes visible the structures and dynamics of interpersonal networks, to understand how people develop and maintain these relations. People may form ‘ties’ of different sorts, ranging from extended, direct interaction reflecting significant ties, to more indirect ties.

- **Discourse Analytics**: Analytics could go beyond simple quantitative logs, and provide feedback to educators and learners on the quality of the contributions. Researchers are beginning to draw on extensive prior work on how tutors mark essays and discussion posts, how spoken and written dialogue shape learning, and how computers can recognize good argumentation, in order to design analytics that can assess the quality of text, with the ultimate goal of scaffolding the higher order thinking and writing that we seek to instill in students.

### 4. Requirements of the Reference Software Architecture

In this section, we present preliminary results of requirement analysis of the reference software architecture based on the survey described in the previous sections, Section 2 and 3. Even though, the requirements are multifold over data, analysis, and application requirement, the results can be summarized as design requirements of reference software architecture as follows:

- **Open and extensible**: It should be open to incorporate new sensors or analytics functionality, desirably without interrupting the task being serviced. It also should ensure incorporation or modification of new workflows at the task level.

- **Distributed**: It should be able to handle multiple sources of data and functionalities distributed over multiple systems. It is also desired to be able to distribute data and to delegate functionality dynamically and transparently.

- **Interoperable**: It should provide compatibility for various learning platforms or VLE by providing interoperable interface to the data and operations.

- **Reusable and configurable**: The functional components and data interfaces should be modular and thus reused and configured for different tasks or more complex tasks as building blocks.

- **Real-time and predictable**: Learning analytics should be performed satisfying the real-time constraints and should be able to estimate the time to completion.

- **Usable**: It should acceptable user experience (UX) by providing appropriate data visualization and user interfaces for monitoring and tasking throughout the learning analytics process.

- **Secure and traceable**: It should protect personal user information to secure privacy and preserve confidential information. Some analytics functionality should be ensured not to be performed as
required. Furthermore, the history of execution of analytics functions and access to data should be recorded, if needed, to ensure traceability.

5. Reference Software Architecture

A reference software architecture is a software architecture where the structures and respective elements and relations provide templates for concrete architectures in a particular domain or in a family of software systems. A reference architecture often consists of a list of functions and some indication of their interfaces (or APIs) and interactions with each other and with functions located outside of the scope of the reference architecture (Wikipedia, 2014b).

In this section, we present our preliminary reference software architecture as shown in Figure 4. The main purpose of this initial work is to identify the necessary components and data interactions of the components before the results of the requirement analysis are fully applied. The initial architecture is then instantiated with supporting software tools to validate the effectiveness. Table 1
summarizes the basic input and output interfaces for each component and applicable software to implement the functionality of the component.

Table 1: Specifications of the Reference Software Architecture

<table>
<thead>
<tr>
<th>Component</th>
<th>Description</th>
<th>Input</th>
<th>Output</th>
<th>Applicable Software</th>
</tr>
</thead>
<tbody>
<tr>
<td>Learning Sensor API</td>
<td>External API to Collecting Learning Activity Data</td>
<td>Metric Profile</td>
<td>Learning Event Data</td>
<td>Apache Storm</td>
</tr>
<tr>
<td>External Database Interface</td>
<td>Collect Structured Data from RDB or Web Services</td>
<td>Well-Formatted Data</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LOD Triple Reasoner</td>
<td>Collect LOD and extension Triple Data</td>
<td>LOD Triple</td>
<td>LOD Triple</td>
<td>Webpie</td>
</tr>
<tr>
<td>Statistical Data Extractor</td>
<td>Calculate Statistical Information</td>
<td>Literal data set</td>
<td>Analyzed Information</td>
<td>R</td>
</tr>
<tr>
<td>Learning Pattern Recognizer</td>
<td>Recognize Specific pattern by Data Mining</td>
<td>Literal Data Set, Analyzed Information</td>
<td></td>
<td>R</td>
</tr>
<tr>
<td>Adaptive Recommender</td>
<td>Recommend Associated Content</td>
<td>Analyzed Information</td>
<td>Literal Data</td>
<td>R</td>
</tr>
<tr>
<td>Visualization Tools</td>
<td>Visualize each Types of Analyzed Data</td>
<td>Analyzed Information</td>
<td>HTML</td>
<td>Google Charts</td>
</tr>
</tbody>
</table>

6. Discussion and Future Work

Realizing learning analytics systems indispensably requires an integrated and holistic approach. In this paper, we first identified the basic requirements of reference software architecture to capture the holistic and integrated view. The primary requirements includes the system to be open, distributed, interoperable, reusable, real-time, usable, and secure. So far these requirements were derived from the survey of the state of the art in learning analytics field, but they need to be refined and applied to the reference software architecture by the real use cases and application scenario. We are currently collecting such use cases and application scenarios of learning analytics. We are well aware that our goal to develop the explicit specifications of the learning analytics architecture is far away from this preliminary work, but we believe this is an essential step toward the goal.

Acknowledgements

This research was supported by the ICT Standardization program of MISP (The Ministry of Science, ICT & Future Planning)

References

Learning Analytics Interoperability – looking for Low-Hanging Fruits

Tore Hoel a* & Weiqin Chen b

aOslo and Akershus University College of Applied Sciences, Norway
bUniversity of Bergen, Norway
*tore.hoel@hioa.no

Abstract: When Learning Analytics is seeking a wide community, the challenge of efficiently and reliably moving data between systems becomes important. This paper gives a summary of the current status of Learning Analytics Interoperability and proposes a framework to help structuring the interoperability work. The model is based on a three dimensional Enterprise Interoperability Framework mapping concerns, interoperability barriers and potential solutions. The paper also introduces the concept of low-hanging fruits in prioritising among solutions. Data gathered from a small group of Norwegian stakeholders are analysed, and a list of potential interoperability issues is presented.

Keywords: Interoperability, Standards, Learning Analytics, Educational Data Mining, Sharing Data Sets

1. Introduction

Learning Analytics (LA) is an emerging research field where we are starting to see contributions from a diversity of research disciplines, and development of a range of tools, techniques and applications used by LA researchers and practitioners (Siemens, 2013). However, large-scale implementations of LA in an educational sector, a region, an industry, – or even in an institution, remain to be seen. Scaling up LA means to go beyond research prototypes or the innovative solution of a single vendor who keeps the data under tight control within a closed ecosystem, the analytics magic in a black box, and only exposes the results to the users in colourful dashboards. Unless scaling up means ‘winner takes all’, we need to address a range of new issues posed by the needs of actors, systems, organisations, and cultures to interoperate.

Data lies at the heart of learning analytics. This does not necessarily mean that data sharing and interoperability has been a main concern for LA research or development till now. Interoperability involves different aspects of how systems at large (both organisations and ICT systems) communicate on different levels (e.g., technical, semantic, organisational, political, and legal). New challenges are posed when scenarios foresee third party LA tools analysing data from diverse sources by national and international organisations sharing and comparing data. In addition, moving from prototypes to large-scale implementations opens up a raft of new issues, – organisational capacity and privacy being only two of them (Scanlon et al., 2013; Siemens, 2013).

We suggest using the concept of interoperability as an overarching term for this new level of discourse on scaling up applications of LA. By doing so, we bring a new set of actors to the table, underlining that user groups, implementers, standardisation experts and bodies, local authorities, and others have a role to play in order to reap the benefits of bringing analytics to education (MacNeill & Campbell, 2014). Interoperability as a term will be discussed below. However, we would also suggest to apply the concept of “low-hanging fruits” in framing the discourse on LA interoperability, as “the way ahead to get results sometime soon requires care (..) a middle way seems necessary, in which a little time is spent on discussing the most promising and the best-understood targets, i.e. to look for the low hanging fruit” (Cooper, 2013b).

This proposal to identify low-hanging fruits addresses the problem of how to conceptualise the solution space for learning analytics. Our background in the standards community has made us wary
of big and “complete” designs never leaving the researchers’ drawing pad. LA have the potential to associate numbers with any aspect of the learning process, and as a consequence requires immensely complex data models for exchange of information. With ambitions of large-scale implementations based on an ill-defined problem space, a piecemeal and lightweight approach might be more advisable (Hoel, 2014a; Sales et al., 2012). Consequently, there is a need to find a way to identify the low-hanging fruits.

In this paper we have carried out a pilot study exploring the Learning Analytics Interoperability (LAI) problem space by interviewing a small number of representatives of LA stakeholders. The rest of this paper is organised as follows: First, the concepts of interoperability and low-hanging fruits are reviewed. Then a small explorative study of stakeholder groups’ views on interoperability in the context of learning analytics is presented. The results are discussed in relation to an Enterprise Interoperability Framework, searching for approaches to interoperability that could be characterised as low-hanging fruits.

1.1 Learning Analytics Interoperability

A search in Google Scholar on ‘learning analytics’ AND ‘interoperability’ gives in mid-2014 just above 400 hits; while searching for ‘learning analytics’ AND ‘data sharing’ gives less than 100 hits. In 2013, Cooper surveyed academic and formal publications as well as informal publications and noted that “only a small group of people, largely researchers, have drawn attention to LAI and a significant amount of the literature has been produced by a few people” (Cooper, 2013a). He also found no references to LAI from software suppliers.

“The way LAI is covered by these works will be identified as being of three kinds: assertion or argument in favour of interoperability in general; references to interoperability for a particular purpose or context; interoperability as a significant or key topic. Assertion and argument about interoperability are usually concerned with the lack of it” (Cooper, 2013a).

There has been little work on interoperability specifications by the educational technology community; first in August 2014 the ISO committee working on learning technology standards established an ad hoc group to develop scope for new work items on LAI.

Interoperability is a multidimensional term with many interpretations and definitions. According to the Institute of Electrical and Electronics Engineers interoperability is “the ability of two or more systems or components to exchange information and to use the information that has been exchanged” (Geraci, 1991). Cooper (2014) states that “a broad interpretation of “systems” that includes people and the activities they undertake using these digital technologies captures the true essence of interoperability as a means to achieve human aims and objectives”. However, without describing the different dimensions of interoperability the term tends to get a merely technical interpretation, leading to a focus on exchange of data when there is a need to zoom out and look at the social, political and organisational motivators and barriers to interoperability. We would suggest that a perspective inspired by Enterprise Interoperability (EI) should be applied at this early stage of exploring LA Interoperability (LAI) challenges. In the EI setting interoperability is defined as the “ability to (1) communicate and exchange information; (2) use the information exchanged; (3) access to functionality of a third system” (Chen & Daclin, 2006).

Applying an enterprise perspective to interoperability foregrounds the two dimensions that make up the problem space (barriers and concerns), and highlights the need to explore the solution space looking into alternative approaches, as illustrated in Figure 1 (Chen & Daclin, 2006). Analysing the barriers it will make sense to group them in the broad categories of conceptual, technological and organisational barriers. The concerns however, need to be derived by studying the domain characteristics of education as a particular instance of an enterprise. The dimensions identified in the ATHENA Interoperability Framework (http://athena.modelbased.net) with concerns related to data, services, processes and business, might help the analysis.
A key task in the analysis of the problem space and the solution space will be to identify LA solutions that address stakeholder concerns and overcome interoperability barriers. When the solution space is mapped, the next step is to choose strategy; and it is here we will introduce the concept of low-hanging fruits.

1.2 Low-Hanging Fruits

By using tools and data that are already in place the community could benefit right away from the development of new knowledge and new designs. This is the basic idea behind the approach of reaping the low-hanging fruits. To extend the fruits metaphor, one should refrain from extensive pruning (e.g., changing the context or the system) until the gardener knows more about the trees and the garden. Within Learning Analytics and Educational Data Mining this may make sense, since it is difficult getting data out of information systems. However, the hunger for tasting the benefits of LA is great; the potential data sources are diverse; and the range of methods and experience is growing (Cooper, 2013b). By going for the low-hanging fruits we allow stakeholders time to argue their case for specific LA solutions before deciding on approaches with far-reaching implications.

2. Related work

In 2011 SOLAR, the Society for Learning Analytics Research, issued a proposal to design, implement and evaluate an open platform to integrate heterogeneous learning analytics techniques under the name of Open LA (Siemens et al., 2011). The proposal was a high level argument in favour of openness of process, algorithms and technologies; and modularized integration, – asking for development of common language for data exchange and open repository of anonymised data. It is still early days to deliver on this proposal; in 2014 a follow up meeting was organised where SOLAR joined forces with the Apereo Foundation, an umbrella organisation for a number of open source projects. Their aim for a LA Initiative is now to “accelerate the operationalization of Learning Analytics software and frameworks, support the validation of analytics pilots across institutions, and to work together so as to avoid duplication” (Cooper, 2014a).

LA Interoperability initiatives are also launched by the standards community. Since 2010 the Advanced Distributed Learning (ADL) initiative have developed an eXperience API (xAPI), also called TinCan API, based on the idea of tracking activity streams (ADL, 2014). A similar approach is adopted by IMS Global in their Caliper project (IMS, 2013), initiated late 2013. When ISO/IEC JTC 1/SC36 late 2014 starts to work on LAI it is assumed that they will begin by defining an abstract framework in what eventually could become a multipart international standard (Hoel, 2014b).

The Open LA and the xAPI initiatives represent the opposite parts of the LAI continuum. While the former is more an interoperability dream, the latter represents a very concrete approach to
exchange data on any activity that is related to learning, storing statements of the form “I did this”, linking an actor to an object via a verb. However, the main parts of the LAI continuum are still to be addressed. Cooper (2014) has reflected on the “big picture” of LAI exploring what should be the scope of work in this field. He identified three areas of discourse defined by these questions:

- **Models and Methods**: How can we transfer information about statistical and data mining methods, the parameters used, and the predictive models?
- **Analytical Results**: How can we transfer individual-level and grouped numerical results? How can we track data provenance, quality and processing methods?
- **Data for Analysis**: How can we get data out of the operational systems? (Cooper, 2014)

One takeaway from Cooper’s briefing (2014) is that LAI is very complex and involves interoperability specifications that are not generally known within the educational technology community. As an example, Cooper mentions PMML, the Predictive Model Markup Language (dealing with interoperability of models and methods); and SDMX, the Statistical Data and Metadata eXchange standard (dealing with interoperability of analytical results). To further LAI one needs to invest in knowledge building to deal with this complexity, which is not only of technical nature, but also have issues concerning consensus about the objectives of LA as a whole.

### 3. Soliciting stakeholders’ interoperability requirements

When Big Data is promoted by global consultancy firms as “the next frontier for innovation, competition and productivity” (Manyika et al., 2011), and research is pointing to large datasets as the key to improving learning and the environments in which it takes place (Ferguson, 2013), it is essential to solicit the views of stakeholders who play vital roles in delivering and analysing the data and using the results. Data sharing and interoperability between IT systems are about tearing down unwanted barriers. It is important to note that this is only one perspective, and that the flip side of barriers could be boundaries people have established to protect themselves. We only know what perspectives will influence design and implementation of LA applications when we have involved the stakeholders.

This paper presents a pilot study aiming to structure the discourse on interoperability and give input to scoping of the first work items for standardisation in this field. We have interviewed eight representatives of students, teachers, support staff, and policy makers in Norway asking them to elaborate on interoperability in the context of LA. The semi-structured interviews focussed on what are the things to agree upon (that need to interoperate) to realise the potential of LA, aims of LA applications, and what data sources could be used for LA. The interviews were supported by an online sticker board, resulting in graphical summaries of the interviews giving a rough sense of the priorities of the respondents.

This research is positioned in the first Relevance Cycle of the three research cycles of Design Science (Hevner, 2007; Hevner et al., 2004), addressing requirements and field testing. The purpose is to come up with candidate concepts that describe the problems and opportunities in the application domain from a people, organisational systems, and technical systems perspective. The study is offered as an approach that could be replicated within other regional or national communities or sectors in order to gather requirements for LAI.

### 4. Results

The first results are an unordered list of concerns of the respondents being asked to reflect upon LA. The interviews are by no means representative for Norwegian communities or for the different stakeholder groups. Nevertheless, this study indicates that it is still early days for the idea of using LA in schools and higher education. There is little experience with LA solutions, and there is a general need for overview and understanding of the basic ideas. Even if the concerns reflected the respondents’ role in education, the issues they brought forward reflected more or less the same views on the problem space. Support for the individual learner, and emphasis on privacy and control of the data generated through interaction with the systems were in the foreground of every interview. The concerns mentioned were:
**Student interviews**

- Personal development and support for learning and career planning as primary aim; helping the institutions, e.g., to improve retention, as a secondary aim
- Non-intrusive guidance (not being evaluated & tested all the time)
- Privacy & Control over personal information
- Trust - school or university as a trusted “partner” in LA
- Consent to allow data flow between systems - transparency to who sees what
- Access to data for LA: Students have a mixed use of tools that do not exchange data
- Educational tools policy: More emphasis on institutional tools, like LMS, will prevent students from using social networking tools for learning. Latter group of tools important for life-long learning.
- Control over LA results: Students want to be in control of interaction with their data - ownership to analysis and results, not only to data
- Data should be open, based on agreements between the partners involved
- Coordination and coherence among services: Students enrolled in different courses, learning on different platforms - in order to get a coherent picture teachers, data, etc. need to be coordinated
- LA solutions should allow for two-way interaction, user control, consent, time to think before giving data away, etc.
- Non-intrusive LA (no extra time on using LA tools)

**Teacher interviews**

- Understanding the affordances of LA: To get a conceptual understanding of the domain, from different perspectives, not only technological aspects
- Prioritising benefits in this order: Individual (adaptive), teaching, institutional / organisational
- Making sense of ‘contexts’ and ‘activities’: One cannot make sense of data, unless one knows their context. How to describe contexts and activities?
- LA implementation without losing control of «pedagogy» (technology or market driven vs. pedagogically motivated innovation)
- Learner Control and ownership to data; and control over how data are used, e.g., through anonymisation / pre-processing (removing personal information)
- User Control over tools & services
- Getting the statistics out of the tools that are currently used
- In listing potential sources of LA data, institutionally controlled data sources, e.g., LMS, are mentioned first

**Support staff interviews**

- Interpretations of LA results: What do we measure - and what does the results mean related to the different stakeholders’ use of the LA results?
- Control over data: Do students have the right to reserve themselves against participating in LA, sharing data from LA, etc.?
- Agreeing about contexts for analytics. What to do with ‘context free data’ that make no sense for analysis?
- Primary LA beneficiaries should be the learner
- Need to improve the interoperability of legacy systems in order to get data for LA
- Need to agree upon realistic aims for the use of LA
- Promote institutional control of data generated in (cross-institutional or international) MOOC systems
- Develop and introduce systems that give enough data to allow LA, e.g., MOOCs
- Start using non-controversial data - e.g., data showing if watched videos are too long, before using data identifying the individual

**Policy maker interviews**
• Creating a culture for LA - mapping the incentives to make use of LA and develop a strategy
• Privacy and Ownership to data - Open Badges approach to data (owning your own data)
• Interpreting data: How to avoid measuring the wrong data and making invalid conjectures?
• The ultimate aim is to get empirical support for pedagogical choices, improve quality of
  learning resources, and further adaptive learning. However, also institutional aims are
  important, e.g., better retention and early warnings of drop outs
• We should discuss who should store the data; should it be nationally controlled or distributed?

Concerns can be grouped along a continuum starting with data in the bottom row, climbing up via
Tools and Technology Support (Services), Learning Activities (Process), and ending with Aims
(Business) at the top row. We have chosen these dimensions as a refinement of the categories used in
Enterprise Interoperability Analysis (Chen & Daclin, 2006). This tradition looks at Conceptual,
Technological and Organisational barriers as the categories for the barrier axis fitting well with the
investigation of our study.

4.1 Problem space

Table 1 presents the results of populating the LA problem space with the data from the stakeholder
interviews. The numbers in the table both refer to the concern/barrier nexus described in the text
below, and give an indication of urgency or priority gleaned from repeated rounds of qualitative
analysis of the interviews by two independent researchers (1 representing the highest
urgency/priority).

<table>
<thead>
<tr>
<th>Concern/Barriers</th>
<th>Conceptual</th>
<th>Technological</th>
<th>Organisational</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aims</td>
<td></td>
<td>(1)</td>
<td></td>
</tr>
<tr>
<td>Learning activities</td>
<td>(3)</td>
<td>(2)</td>
<td></td>
</tr>
<tr>
<td>Tools and Technology support</td>
<td></td>
<td>(5)</td>
<td></td>
</tr>
<tr>
<td>Data</td>
<td></td>
<td>(4)</td>
<td></td>
</tr>
</tbody>
</table>

(1) Privacy, Trust & Control of Data: This nexus between enterprise aims and organisational
barriers relates to the complex issues of how interoperating LA systems get access to data without
violating the privacy of users; and how to maintain legitimacy of these systems while giving the users
control of their data.

This problem is situated in the most abstract corner of this two dimensional space. It relates barriers to
interoperability on all levels. However, there is no conceptual or technological fix to this problem; it
is clearly up to organisations and their members to agree upon questions like how much private
information has to be exposed in order to reap the benefits of more adaptive systems, more support to
learning, etc.; what institutions within education can be trusted to manage personal information for
which groups of learners, with what kind of procedures; and what kind of control will the system give
the originator of data throughout the LA cycle. And perhaps most importantly, what aims should have
priority, e.g., if there are conflicts between the aim of an institution to reduce drop out and the privacy
of the student?

(2) LA affordances and application domains: This interoperability problem arises when there is no
consensus about the benefits of LA, and what domains LA should be applied to.

This problem is related to strategies for policy development and implementation for institutions,
sectors and governments. Even if the barriers are organisational, at this early stage of LA conceptual
barriers (e.g., lack of shared vocabularies) are part of this problem.
(3) **LA Context & Learning Activities**: This is the “blind data” problem that arises when there is no context information provided with data from a learning activity.

This is a conceptual barrier due to the lack of linkage between learning activity streams and their pedagogical contexts.

(4) **Legacy system interoperability - information model for LA data exchange**: This is the classic learning tools interoperability problem where systems have data in silos without any possibility of aggregating data to get a coherent view of the activities in a class or a school.

When LA is added as yet another system this data integration problem is brought to the attention again.

(5) **LA implementation best practice guide**: This interoperability problem relates to the market vs. policy driven implementation of LA systems and lack of open institutional or regional LA policies.

Technical interoperability is not always first priority for an enterprise developing an innovative LA solution. Organisations may find they need guidance to best practice for implementing LA solutions in order to support other educational policies.

4.2 Searching for solutions and low-hanging fruits

Having mapped the problem space, the next step is to develop the solution space, adding a new approach dimension to our model. As explained in Figure 1, a Solution Space is formed when Approach, Barrier, and Concern intersect. At this stage we are not searching for any solution but the one that can be picked as a low-hanging fruit.

Our interviews only indirectly pointed to solutions. However, the results reported in Table 1 give a prioritisation of problems, which is a first step towards designing an approach. Our data clearly shows that the non-technical issues are seen as the most imminent and important problems that could jeopardise uptake of learning analytics. The barriers are mainly organisational and conceptual, with the only identified technological problem being related to legacy systems.

The following approaches are based on a second analysis of our data to extract possible ideas for solutions.

**Privacy, Trust & Control of Data**

Strictly speaking, this complex set of issues is not related to LA in particular. Some experts may even say it is out of scope for LAI. Our respondents, however, see this issue as major stumbling block that needs to be dealt with, even before discussing the potential benefits of innovative LA solutions now being marketed. It is also an issue that is manageable, at least from a conceptual and technological point of view. The challenge is to engage in this work a new group of stakeholders (e.g., teachers, policy makers) whose primary interest is not with technology development.

Privacy laws are in place for schools and universities. Too strict interpretations may stifle new use of technologies and new learning practices as seen in reluctance to use non-institutional controlled services (e.g., cloud services). There is a need to clarify rules and practices (e.g., lay down principles for which services to trust). This may prove difficult, as the new learning practices that push the boundaries away from institutionally controlled, teacher-led education towards learner-centred, socially situated life-long learning may challenge existing privacy and data protection paradigms.

Our respondents pointed to learner control over data and teacher control over pedagogy as principles to pursue. We will come back to this point in the Discussion section of this paper.

**LA affordances and application domains**

Our respondents see the benefits of LA, but have a lot of questions mixed with fears about potential adverse effects. Are we measuring the right things? Do we discern between causation and correlation? Are we able to draw the right conclusions from the LA dashboards? To improve interoperability between LA systems one needs a more granular picture of LA activities. Academic analytics to prevent dropouts from school is a different activity to Learning analytics capabilities built into a digital textbook to support adaptive learning. As one respondent put it, “if we only have data for when
a video loses the audience (not knowing it was Peter), we should focus on improving the video, not
dreaming of supporting a particular person”.

The interviews requested more conceptual clarity to the field of LA. This is a feasible task that
could be organised as a consensus process.

**LA Context & Learning Activities**

The approach to solve the problem of missing context descriptions in relation to learning activities is
to launch a traditional standardisation project developing the necessary vocabularies and find ways to
describe relations among the concepts. These are continuous activities within learning technology
development. However, the new interest in LA may contribute to speed up this activity and bring new
stakeholders on board.

**Legacy system interoperability - information model for LA data exchange**

LA could potentially be a driver for revisiting the interoperability problems due to siloed legacy
systems. However, as one respondent said, nobody is going to rebuild the student information system
that is working well for all higher education institutions in Norway. The solution may be in
establishing some kind of aggregated system, raising the question of who should run such a system,
and how should the data be stored. Again, in order to make progress on this problem, the first problem
cluster discussed above needs to be sorted out.

**LA implementation best practice guide**

To lead by example is a good principle. A best practice guide may help institutions to implement LA,
as such guides need to be clear about what the actors of the implementations should be. Developing a
guide as a consensus document may help the different levels of the educational system to identify
other interoperability problems and to prioritise among them. Starting in the nontechnical end of
interoperability work may first, help address the concerns of the stakeholder groups represented in our
survey; and second, take the focus away from technical LA challenges, making it easier to address the
semantic, organisational, political and legal interoperability problems posed by putting numbers in the
service of learning, education and training.

5. Discussion

Traditionally, ISO/IEC JTC 1/SC36 had published a number of multipart standards within the
learning technology domain. When launching an LA project, this international standardisation group
would be expected to fall into the old pattern starting with a first part being a LA Framework
standard, and the other parts filling the different puzzles of an all encompassing LA jigzaw. This
paper can be read as a warning against such an approach. The explorative interviews we have reported
make it clear that detailed information models alone will not ease the uptake of learning analytics in
schools and higher education. In order to foster interoperability among actors in this sector there is a
need to find solutions on all levels where two systems interoperate. The big question, however, is to
find which puzzle to start with that will make it easier to see the pattern and find the solution to the
other pieces.

We have suggested designing a solution space by soliciting input from stakeholders that do not
necessarily know much about LA, but who eventually will play a crucial role in its adoption. In this
study, the respondents highlighted above all the softer issues related to privacy, trust and control. It is
worth exploring if the solution space related to this issue could be established as a kind of baseline for
further design of LA systems.

Often the Big Data hype is used to sell Learning Analytics (Ferguson, 2012), triggering stories
of aggressive and subtle marketing and manipulation from a commercial context. LA needs to
distance itself from the setting of marketing and sales, as learning and education have very little in
common with the motivational arm-twisting of commerce. From a liberal, and some may say Western
perspective, the ethos of learning and education is that the learner should be in control and the
supporting institutions should only do what is in the interest of the learner. Therefore, it should be
easy to argue for learner control over her own data; transparency controlled by the individual learner;
trust built bottom-up; etc. The counterargument would be that such an approach will not give the
amount of data needed for LA, as only centrally run systems where all are subscribed would give enough data. However, in a civic society the individual has some rights to opt out of education. And privacy protection and trust are not opposites. Trust does not need to be blind; it can be a dynamic property, ultimately controlled by the learner, but also maintained by institutions, e.g., the school, the university or the educational authorities.

What happens if the learners (and their parents) lose trust in LA systems is vividly illustrated by the US InBloom case:

Protests began in earnest when it was discovered that InBloom’s software had more than 400 optional data fields that schools could fill out—asking for potentially sensitive information such as the nature of family relationships, learning disabilities, and even Social Security numbers. Although there were no reported leaks, parents were uncomfortable without an absolute guarantee of that data’s safety or a clear indication of who could access it. (Slate future tense, 24 April 2014)

InBloom “was meant to extract student data from disparate school grading and attendance databases, store it in the cloud and funnel it to dashboards where teachers might more effectively track the progress of individual students” (New York Times, 21 April 2014). In April 2014, after a period of heated public debate, the system had to close down after the New York state passed legislation prohibiting the state department of education from giving student information to data aggregators like InBloom.

Giving priority to the solution of privacy, trust and control issues could help identifying the LA systems with the best possibilities to succeed within the nearest timeframe. Trust is built in concentric movements starting with the learners, co-learners, school, other community of learners, etc. Local, distributed, transparent and adaptive systems supporting the learner seem to be easier to sustain than systems that are more distant and leave the user with more questions of who is in control. On the other hand, we know that learners use cloud services and social media systems where they have minimal control.

If the systems are found useful, they tend to be used and the users freely give access to their data. In the case of social media, however, the educational institutions are not acting as intermediators between the users and the systems. In formal education, institutions have to follow standards, and it is therefore problematic to mandate use of tools with poor or unknown data protection policies. It would therefore be helpful to have a consensus about how tools outside institutional control are used and what privacy, trust and control models education would promote (Slade & Pinsloo, 2013). This is an argument for engaging in a process on clarifying LA affordances and application domains, the second approach coming out of our stakeholder survey.

Establishing learner control as a design baseline would help identify which LA contexts that need specification. One might assume that smaller systems that are able to demonstrate benefits to the learner would be easier to introduce, and as such represent the low-hanging fruits of learning analytics. While ideas about more complex and institutionally motivated systems, e.g., with institutional, regional or even national learning record stores, should be left to ripen before brought to standardisation. It is also reasonable to think that once the idea of complex and integrated systems are put on the back-burner, new ideas could be foreseen how analytics can be carried out to improve education with existing data and systems. In an emergent field there is a need to showcase and demonstrate best practices that work before investing too heavily in wild dreams.

6. Reflections and Outlook

This research was conducted in accordance with design science guidelines (Hevner et al., 2004) to develop a support framework for structuring work on LAI. Stakeholders were interviewed about their concerns about LA in order to construct a problem space along the dimensions of concerns and interoperability barriers. The interview data was then analysed in order to identify approaches that could be dealt with within a reasonable timeframe, given the dynamic nature of current learning analytics development. The analysis gave five candidate issues that are potential new work items for LAI standardisation.
The results of this study need to be validated through further field-testing in order to see if the same issues are prioritised by more representative selections of stakeholders and respondents from other countries. Furthermore, the relevance of the barriers used in this study should be tested. It is possible that a more fine-grained categorisation may be needed, especially to understand the technological barriers to LAI.

The concept of low-hanging fruits should also be further developed. In this paper we have used the concept to add a strategic dimension to the approach axis of the solution space. We have also indicated that selecting the low-hanging fruits may alter the space itself, giving priority to a certain group of applications, repurposing others. In developing this framework related existing work on quality models for standards looking into aspects of product quality, process quality, and quality in practice (Folmer, 2011) should be explored.

References


Ferguson, R. (2013). Learning Analytics for Open and Distance Education, 1–8.


Making Sense of Online Learning Behavior: A Research on Learning Styles and Collaborative Learning Data

Meng SUN* a, Jiu-Tong LUO* a, Dong-Ming QIANb & Xiao-Qing GU∗ a*

a Department of Educational Information Technology, East China Normal University, China
b e-Educational System Engineering Research Center & Shanghai Engineering Research Center of Digital Education Equipment, East China Normal University, China
∗xqgu@ses.ecnu.edu.cn

Abstract: This study focused on the relationship between learning styles, online behaviors and group collaborations. Sixty junior students from a university in China were taken as research object. Index of Learning Styles was used as a measuring tool to test participants’ learning styles. Relationships between variables were measured by using bi-variate correlations analysis and one-way analysis of variance respectively. The results revealed a meaningful relationship between learning styles and online collaborative behavior. In addition, groups’ online collaborative performances could be significantly different. However, grouping by learning styles might not be the factor that make effects on group collaborations.

Keywords: Learning styles, online learning, online behaviors, group collaborations

1. Introduction

The emergence of “big data” in education holds promise for improving learning processes in formal education, and beyond as well (Siemens & Baker, 2012). Currently, the learning management system (LMS) has been widely used, and has stored a lot of data, which notably supporting researchers’ studies. The analysis of learners’ online behaviors is a research focus in educational technology area. In the field of cooperative learning, what factors can affect groups’ performances is also a topic worthy to explore.

It has been proven that learners’ personality can affect learner behaviors in online learning environment. Many educators consider learning styles as an important factor to influence students’ learning process. Furthermore, education research has emphasized that collaborative learning could improve project quality and performance (Soliman & Okba, 2006). The way in which students are grouped may affect the group performance. One of the features that can be taken into consideration when grouping is students’ learning styles.

The aim of this study was to examine the relationship between learning styles, online behaviors and group collaborations. For this purpose, the following research questions will be answered:

- Which dimensions of learning style have effect on learners’ online behaviors?
- Which kinds of online behaviors could be affected by learning style?
- Is there a significant difference among groups’ online performances?
- Is there a significant relationship between group members’ learning styles and groups’ online collaborative performances?

2. Theoretical framework

LMS such as Blackboard, Moodle, Sakai et al., and many other learning systems have been widely used in e-learning around the world, providing teachers and students with a great variety of features which can be included in the course such as learning material, quizzes, discussion forums, assignments, and so on (Graf & Liu, 2008). With a lot of e-learning behavioral data produced, learning analytics (LA) has
emerged as a new technology aimed to make sense of these data. Learning analytics is a technology focused on measurement, collection, analysis and reporting data about learners and contexts, for purposes of understanding and optimizing learning and the environments in which it occurs (Siemens et al., 2011; Siemens, 2012). Research on learning analytics has found that learning behavior data can predict students’ learning to some extent.

Learning style is a unique combination of primary forms of processing information as well as the way in which various techniques and personal idiosyncrasies are used, and it has been thought as one of the factors which can affect person’s behavior. Research on learning style started from 20th century, and many teachers used this theory while they were teaching. There are various definitions and classifications of learning style.

Learning style is the way in which each person absorbs and retains information and/or skills, regardless of how that process is described, it is dramatically different from each person (Dunn, 1984). Pask (1988) defined learning style as a kind of strategy that learners like to use when they were processing a specific information. And Kolb (1999), Lotas (1977) and Oxford (1993) et al. also put forward their own classifications of learning style from different aspects. Soloman & Felder (1997) developed an index of learning styles and was widely used today. In this instrument, learning styles were divided into four dimension, such as sensing or intuitive, visual or verbal, active or reflective and sequential or global.

3. Methods

3.1 The Participants and Context

72 (14 male, 58 female) junior students majoring in educational technology from a university in China took part in the research. They were all in the course on Instructional System Design which lasted for 20 weeks began in March and ended in June. The course task for the students was to work in groups of six on a project assignment. Students could choose their partners freely. During the course, they should do theme discussion at the course platform (Sakai) and upload their homework to the platform. In addition, every group member should discuss about their group projects after class via the instant communication tool (QQ).

3.2 Measuring Tools

This study used the Chinese version of Index of Learning Styles (ILS) developed by Soloman & Felder (1997), and it had been tested in many researches in China. This questionnaire contains 44 items divided into four dimensions: active-reflective, sensing-intuitive, visual-verbal and sequential-global. Each dimension has associated with 11 forced-choice items with each option (a or b). Taking active or reflective as an example, for statistical analysis, it used ‘b’ responses minus ‘a’ responses, and then the ranging of number will between -11 to +11 (Felder et al., 2005). In this study, +11, +9, +7, +5, +3, +1 represented the learning style were active, and -11, -9, -7, -5, -3, -1 represented the learning style were active on the active-reflective dimension. The higher the value, the stronger is the preference. And the rest three dimensions were following the same distinguishing method mentioned above.

4. Data sources and analysis

In the study, results of students’ learning styles inventory and online behavior, including clickstreams of students’ participating in platform activities and groups’ chat logs after class were all expected to be recorded and analyzed. However, one of the 72 student did not finish the questions of the ILS, and one group did not submit their group chat logs. Finally, data from 10 groups (60 students) were evaluated in total. The population was composed of 83.3% female (N = 50) and 16.7% male (N = 10).

Adopting a quantitative method, SPSS 19.0 was used to analyze the data. In order to answers the research questions, bi-variate correlations analysis and one-way analysis of variance (ANOVA) test...
were used; other basically and descriptively statistical analysis techniques such as frequency, percentage and standard deviation of the distribution were also employed.

5. Results

5.1 Students’ learning styles

The average scores of the four dimensions, as shown in Table 1, revealed that the leaning style of all the 60 students were a little more active, sensing, and sequential, but much more visual.

Table 1: The results of students’ learning styles

<table>
<thead>
<tr>
<th>Learning Styles</th>
<th>N</th>
<th>Mean</th>
<th>SD</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>active-reflective</td>
<td>60</td>
<td>0.33</td>
<td>3.77</td>
<td>-9</td>
<td>9</td>
</tr>
<tr>
<td>sensing-intuitive</td>
<td>60</td>
<td>1.10</td>
<td>4.11</td>
<td>-9</td>
<td>7</td>
</tr>
<tr>
<td>visual-verbal</td>
<td>60</td>
<td>5.67</td>
<td>3.73</td>
<td>-5</td>
<td>11</td>
</tr>
<tr>
<td>sequential-global</td>
<td>60</td>
<td>0.17</td>
<td>4.27</td>
<td>-7</td>
<td>11</td>
</tr>
</tbody>
</table>

5.2 The relationship between learning styles and online behaviors

To investigate the relationship between learning styles and online behavior, correlation analysis was performed. One kind of the online behavior was frequencies of participating in platform activities (FA) which mainly consisted of uploading, downloading and posting. The other was average frequencies of speech (FS) in all the six times of after-class group discussion via QQ.

The Table 2 showed that there was significant positive correlation between active-reflective dimension and TS (r = 0.27, p < 0.05). But the other dimensions of learning styles showed no significant correlations.

Table 2: Correlations analysis results of relationship between learning styles and online behaviors

<table>
<thead>
<tr>
<th>variables</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.active-reflective</td>
<td></td>
<td>1</td>
<td>-0.14</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.sensing-intuitive</td>
<td>-0.14</td>
<td></td>
<td>0.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.visual-verbal</td>
<td>-0.20</td>
<td>0.33**</td>
<td></td>
<td>-0.41</td>
<td></td>
</tr>
<tr>
<td>4.sequential-global</td>
<td>-0.20</td>
<td>0.33**</td>
<td>-0.41</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5.FA</td>
<td>0.43</td>
<td>0.15</td>
<td>-0.13</td>
<td>-0.04</td>
<td></td>
</tr>
<tr>
<td>6.FS</td>
<td>0.27*</td>
<td>-0.01</td>
<td>-1.00</td>
<td>-0.07</td>
<td>0.45**</td>
</tr>
</tbody>
</table>

** Correlation is significant at the 0.01 level;
* Correlation is significant at the 0.05 level.

5.3 Differences among groups’ online performances

The online behavior differences among groups were analyzed in one-way ANOVA (see Table 3). It showed that among different groups, students’ collaborative behaviors were significantly different (F=3.260, p<0.01). Students in Group2, 8 and 9 talked much, but students in Group 1, 11 and 12 talked little. However, significant differences were not found among groups at the level of platform activities (F=0.694, P>0.05), which was relatively independent.

Table 3: ANOVA results of groups’ online performances

<table>
<thead>
<tr>
<th></th>
<th>Sum of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>Post Hoc Tests</th>
</tr>
</thead>
<tbody>
<tr>
<td>FS Between Groups</td>
<td>9.010</td>
<td>9</td>
<td>1.001</td>
<td>3.260**</td>
<td>G2&gt;G1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>G8&gt;G1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>G8&gt;G11</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>G8&gt;G12</td>
</tr>
</tbody>
</table>
5.4 **Difference of group members’ learning styles**

It was clear that active-reflective learning style could make a significant influence on one’s speech times, which were significantly different among groups. Thus, it should be analyzed that if it was the group formations that resulted in the difference, and whether the groups with more speech times had more active members or their group members got much higher cores at active-reflective dimension. Under the condition that each group contained both active and reflective members (see Table 4), one-way ANOVA was used to investigate if there was a significant difference of group formation at active-reflective dimension. It could be seen from Table 5 that there was no significant group formation difference (F=1.436, p>0.05). All groups were heterogeneous and group formation were alike, but interaction within groups were significantly different (see Table 3).

**Table 4: Group formation on active-reflective learning style**

<table>
<thead>
<tr>
<th></th>
<th>G1</th>
<th>G2</th>
<th>G3</th>
<th>G4</th>
<th>G5</th>
<th>G6</th>
<th>G7</th>
<th>G8</th>
<th>G9</th>
<th>G10</th>
<th>G11</th>
<th>G12</th>
</tr>
</thead>
<tbody>
<tr>
<td>active</td>
<td>N</td>
<td>4</td>
<td>4</td>
<td>1</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>2</td>
</tr>
<tr>
<td>reflect</td>
<td>N</td>
<td>2</td>
<td>2</td>
<td>5</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>Total</td>
<td>N</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>6</td>
</tr>
</tbody>
</table>

**Table 5: ANOVA results of group members’ learning styles**

<table>
<thead>
<tr>
<th></th>
<th>Sum of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>active-reflective</td>
<td>Between Groups</td>
<td>172.000</td>
<td>9</td>
<td>19.111</td>
</tr>
<tr>
<td></td>
<td>Within Groups</td>
<td>665.333</td>
<td>50</td>
<td>13.307</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>837.333</td>
<td>59</td>
<td></td>
</tr>
</tbody>
</table>

6. **Discussion**

This study explored a new and important issue on the relationship between learning styles, online behaviors, and group collaborations. It implemented a quantitative research methodology to analyze:

- Which dimensions of learning style that has effect on learners’ online behaviors.
- Which kinds of online behaviors that could be affected by learning styles.
- Whether there is a significant difference among groups’ online performances.
- Whether there is a significant relationship between groups’ formation and groups’ online collaborative performances.

This study highlighted the emergent themes such as learning analytics (LA), group collaborative in online environment. And this innovative research also got some meaningful findings from the perspective of LA. The main conclusions were listed as following. Firstly, learning styles could influence individuals’ online collaborative learning behaviors rather than the non-collaborative learning behaviors. Among the four dimensions of learning styles, active-reflective dimension played a key role. This study showed that active learners talked much more than reflective learners in a group, and the more active they were, the more frequently the talked. The findings were found to be theoretically consistent with the predictions of the Felder-Silverman learning style model (Felder & Silverman, 1988). Active learners prefer to process information actively by doing something with the learned material, for example discussing, explaining, or testing it. On the other hand, reflective learners prefer to think about the material and work alone. Regarding group discussing, active learners are expected to post more often in order to ask, discuss, and explain something, while reflective learners are supposed to prefer to participate passively by rarely expressing themselves. As a result, active learners
are fit for group work, and could be advised to play the leading role in the group discussion to create positive and active environment.

Secondly, some kinds of online behavior such as uploading and downloading activities were not influenced by learning styles. Students with different learning styles showed no significant difference in terms of participating frequencies. Such behaviors possibly affected by other learner characteristics like gender or motivation types (Lim & Kim, 2003).

Thirdly, group formations of heterogeneous learning styles might not the main factor that affect online group collaborative performances. All groups contained both active and reflective learner, but their averages speech times were significantly different. In other words, active learners in some groups spoke more than which in other groups. It is similar with the reflective learners. Some researches maintain that people in heterogeneous groups have the opportunity to learn from each other and will have better outcomes. Students believe that heterogeneity has potential benefits for their group performance (Herrmann, 1987; Kyprianidou et al., 2012). It was interesting in this study that heterogeneous groups could also perform differently. Other factors rather than learning styles might play a key role, like “group members’ reliability, good mood, mutual respect and empathy, clearly agreed goals, willingness to help, flexibility and adaptability” (Kyprianidou et al., 2012).

7. Conclusions and future work

In this paper, the relationship between learning styles, online behaviors, and group collaborations are addressed. Students with different learning style preferences showed significantly different online behaviors in some patterns. These results seem to be important when instructors design course contents and set course tasks. However, not all group collaboration performances could be affected by learning styles. For example, group discussion could be rather different while the group formation were alike in terms of learning styles. As a result, instructors should consider more factors when grouping the students.

Future work will deal with correlations of learning styles and other online behaviors such as time spent on examples, exercises, self-assessment tests, content objects and so on. The Sakai platform should be re-developed to do the analytic work and give the students more suggestion on their study. Moreover, qualitative analysis methods should be used for group discussion contents in order to explore what are the main factors that affect non-collaborative behaviors on the platform and group performance.

References

Herrmann, N. (1987). Creativity, learning, and the specialized brain in the context of education for gifted and talented children. Adapted from an address to the 7th world conference on gifted and talented children, Salt Lake City, Utah


How can Learning Analytics fit into a General Evaluation Framework and already be addressed during Learning Design?

Christian M. STRACKE*

aTELIT Research Institute, University of Duisburg-Essen (UDE), Germany
bEast China Normal University (ECNU) in Shanghai, China
*christian.stracke@uni-due.de

Abstract: In this paper, the author describes how learning analytics can be included within a general evaluation framework and be defined from the beginning of learning design for a learning opportunity. It is analyzed what part of a general evaluation framework will be covered by learning analytics and how learning analytics can contribute to the impact assessment using the generic Evaluation Framework for Impact Assessment (EFI) as an example. It will thereby be discussed how learning analytics can be addressed at the start of the design phase of the learning opportunity based on the reference process model from the international quality standard ISO/IEC 19796-1. Finally it is indicated how an extension of the learning design specification like PAS 1032-2 can be helpful for the introduction and support of learning analytics and summarized which further research is required.

Keywords: Learning Analytics, Evaluation Framework, Learning Design, Impact Assessment, Impact Measurement

1. Introduction

Learning analytics is becoming more and more a hot topic and important question for organizations and policy makers: How to monitor learning processes and to measure the learning outcomes and results and their impact. Learning analytics are starting to be broadly applied today and raise many open questions and issues concerning privacy and data protection. The different legal situations in all countries and the lack of global agreements are current barriers and the public discussions are only beginning now. All these important issues cannot be discussed here as this paper focuses on the potential integration of learning analytics into a general evaluation framework to address learning analytics already in the learning design.

Thus, in the second section of this paper it is described first how learning analytics can be included within a general evaluation framework: It is analyzed what part of a general evaluation framework will be covered by learning analytics and how learning analytics can contribute to the impact assessment using the generic Evaluation Framework for Impact Assessment (EFI) as an example. In the third section, it is thereby discussed how learning analytics can be included in and addressed at the start of the design phase of the learning opportunity based on the reference process model from the international quality standard ISO/IEC 19796-1. An extension of the learning design specification can be helpful for the introduction and support of learning analytics. Finally a summary is given in the conclusions with a foresight for future research.

2. Learning Analytics in a General Evaluation Framework

In this section of this paper it is discussed how learning analytics can be included within a general evaluation framework: It is analyzed what part of a general evaluation framework will be covered by learning analytics and how learning analytics can contribute to the impact assessment. The guiding question is what relation exists between learning analytics and a general evaluation approach following
the philosophy of Total Quality Management (TQM). For that, the generic Evaluation Framework for Impact Assessment (EFI) will be used as an example that will be introduced in brief first.

2.1 The generic Evaluation Framework for Impact Assessment (EFI)

The Evaluation Framework for Impact Measurement EFI was developed to close a gap for assessing and optimizing the holistic total quality development within learning, education and training. It combines the traditional (internal) evaluation of the processes and developed products with the (external) evaluation concerning the strategic objectives and impact that is becoming more and more crucial due to economic cost pressures and international competition. Through this connection, the Evaluation Framework for Impact Measurement EFI offers an adaptable model for the definition and specification of indicators for both, the internal lifecycle and the external relations.

The Evaluation Framework for Impact Measurement EFI is combining the measurement of two dimensions:

1. (Internal) Impact of (direct) Results as outputs and
2. (External) Impact of Outcomes as indirect results.

Using the Evaluation Framework for Impact Measurement EFI, the following theoretical procedure has to be applied in general:

First, the impact of the internal development and output as direct results will be measured by operative indicators. Within one given project or process the operative indicators will be related to the planned products of the project or process. The measurement of the operative indicators has to focus the two dimensions of the pilot implementations: (1) the internal processes and activities and (2) the internal results (to be tested).

Second, the impact of the external relations and outcomes as indirect results will be measured by strategic indicators. They will be related to the strategic objectives of a given project or process: The measurement of the strategic indicators has to focus the two dimensions of the given project or process: (1) the external processes and activities (within the whole organization and external relations) and (2) the outcomes and their impact and external relations.

The following figure presents the overview of the Evaluation Framework for Impact Measurement EFI and demonstrates its relations between the two dimensions of impact measurement (internal impact of pilots assessed by the operative indicators and external impact of outcomes assesses by the strategic objectives):
2.2 Learning Analytics within the Evaluation Framework for Impact Assessment (EFI)

Learning Analytics is covering a broad range of different processes and results: It can start with monitoring the learning processes and providing feedback to the users as automatic recommendations for further progress as well as to the teachers, tutors or trainers as automatic indication which learners are progressing slowly or even failing and may require specific support and attention. Main objective of learning analytics is the measurement of the learning outcomes and results: Through their analysis it is expected that learning analytics are also contributing to the assessment of the impact. On the other hand, general evaluation frameworks have been developed during the last decades following the philosophy of Total Quality Management (TQM) and dealing with impact assessment, too. Thus, it is important to clarify their relationship and to combine and integrate them.

Therefore the guiding question is what relation exists between learning analytics and a general evaluation approach. General evaluation frameworks for a total quality management are addressing and covering all processes starting with the needs analysis whereas learning analytics is only starting with the learning process itself. Even though the concept and design of learning analytics should be discussed and defined from the beginning, the focus and scope of learning analytics is limited compared with holistic total quality management. Thus, learning analytics has to be part of a broader general evaluation framework like the Evaluation Framework for Impact Assessment (EFI).

Within EFI, learning analytics can directly support the impact assessment of the internal results, i.e. the learning outcomes of the learners achieved within the learning processes of the provided learning opportunity (e.g. by measuring the increase of knowledge, skills and competences in relation to the defined learning objectives). In addition learning analytics can also focus the assessment of the internal processes, i.e. the learning processes (e.g. by assessing the given answers and providing recommendations for further reading and in-depth learning).

For this purpose it is important that the operative indicators defined within EFI for the internal results and processes are also reflecting and aligned with the objectives of the learning analytics. On the other hand the operative indicators for the learning analytics has to be defined carefully that they can also contribute as input for the external impact assessment realized within EFI through the strategic objectives and related indicators. Then the indicators for the learning analytics can not only serve to
monitor the learning processes (as internal products) and to measure the learning outcomes (as internal products) but also to support the assessment by EFI of the external impact on the organization, external stakeholders and the society.

3. Learning Analytics within Learning Design from the beginning

In this section, it is discussed how learning analytics can be addressed at the start of the design phase of the learning opportunity. The reference process model from the international quality standard ISO/IEC 19796-1 is an approach developed and approved in consensus as well as implemented worldwide. As it is covering the full life cycle of any learning opportunity, learning analytics is included even though it is not explicitly mentioned. In the following the role of learning analytics within this general reference process model should be identified: Therefore the international quality standard ISO/IEC 19796-1 will be introduced in brief first.

3.1 The international quality standard ISO/IEC 19796-1

The standard ISO/IEC 19796-1 is the first international quality standard for learning, education and training and provides a common reference framework for learning processes. It was developed in consensus by the Working Group 5 "Quality Assurance and Descriptive Frameworks" of the standardisation committee ISO/IEC JTC1 SC36 and issued by the International Organization for Standardization (ISO) in 2005. It contains the reference process model "Reference Framework for the Description of Quality Approaches" (RFDQ) to support stakeholders in learning, education, and training to document and (re-)define their daily business and processes. The reference process model of ISO/IEC 19796-1 is the integration of the following two main reference models (cf. ISO/IEC 2005):

- the generic process model and
- the generic descriptive model.

The reference process model covers the whole lifecycle of learning, education, and training in general including e-Learning and blended learning. Therefore it can be used to describe any learning scenarios as well as any educational and vocational training product and learning solution. It is important to note that the reference process model does not include any regulations about the sequence of the processes or interdependencies between them as well as it does not give any instructions on its specific implementation in detail as a prescription or regulation. The reference process model serves as an open descriptive framework that always needs the adaptation to the organisation, the learning context, and the given situation. The reference process model is based on the generic process model that is divided into seven process categories containing in total 38 processes. It is described by the following table:
### Table 1: The process model of ISO/IEC 19796-1

<table>
<thead>
<tr>
<th>ID</th>
<th>Category</th>
<th>Description</th>
<th>Processes</th>
</tr>
</thead>
</table>
| NA | Needs Analysis | Identification and description of requirements, demands, and constraints of an educational project | NA.1 Initiation  
NA.2 Stakeholder Identification  
NA.3 Definition of objectives  
NA.4 Demand analysis |
| FA | Framework Analysis | Identification of the framework and the context of an educational process | FA.1 Analysis of the external context  
FA.2 Analysis of staff resources  
FA.3 Analysis of target groups  
FA.4 Analysis of the institutional and organisational context  
FA.5 Time and budget planning  
FA.6 Environment analysis |
| CD | Conception / Design | Conception and Design of an educational process | CD.1 Learning objectives  
CD.2 Concept for contents  
CD.3 Didactical concept / methods  
CD.4 Roles and activities  
CD.5 Organisational concept  
CD.6 Technical concept  
CD.7 Concept for media and interaction design  
CD.8 Media concept  
CD.9 Communication concept  
CD.10 Concept for tests and evaluation  
CD.11 Concept for maintenance |
| DP | Development / Production | Realization of concepts | DP.1 Content realization  
DP.2 Design realization  
DP.3 Media realization  
DP.4 Technical realization  
DP.5 Maintenance |
| IM | Implementation | Description of the implementation of technological components | IM.1 Testing of learning resources  
IM.2 Adaptation of learning resources  
IM.3 Activation of learning resources  
IM.4 Organisation of use  
IM.5 Technical infrastructure |
| LP | Learning Process | Realization and use of the learning process | LP.1 Administration  
LP.2 Activities  
LP.3 Review of competency levels |
| EO | Evaluation / Optimization | Description of the evaluation methods, principles, and procedures | EO.1 Planning  
EO.2 Realization  
EO.3 Analysis  
EO.4 Optimization/ Improvement |

### 3.2 Learning Analytics within the Reference Process Model

The reference process model of ISO/IEC 19796-1 is a valuable and general instrument for the implementation and establishment of quality development in Learning, Education and Training (LET) and beneficial for the introduction of total quality management (cf. Stracke 2010). It has to be identified the role that learning analytics can play within it.
Several processes of the reference process model of ISO/IEC 19796-1 can be identified that are directly relevant for learning analytics and addressed by it: Learning analytics is measured during the process Activities (LP.2) of the process category Learning Process (LP) as part of the process Realization (EO.2) of the process category Evaluation/Optimization (EO). It has to be defined during the process planning (EO.1) of the same process category Evaluation/Optimization (EO) and is finally contributing to the process Analysis (EO.3). As already mentioned above, learning analytics should also be addressed from the beginning of the needs analysis and learning design: Thus, it would be necessary to include learning analytics into the definition of learning opportunities and their needs and design.

The identified processes from the reference process model of ISO/IEC 19796-1 can be used for the further refinement of indicators for learning analytics: As mentioned before, it is crucial that learning analytics are contributing to the general evaluation framework and that the indicators for assessing the impact of the learning opportunities have to be defined already during the learning design process in line with the overall evaluation as well as with the learning analytics.

An extension of the Learning Design (LD) specification developed by Rob Koper and his team at the Open University of the Netherlands in the year 2001 could be helpful for the introduction and support of learning analytics: The open Publicly Available Specification (PAS) DIN 1032-2 was developed by a working group of the German Standardization Body DIN based on the Learning Design: The main amendment is the introduction of the category context to define the environment and its conditions. This additional category is important for the learning analytics as it is providing the basic information for the definition of indicators measuring the learning outcomes.

Further research can reveal and transfer these conditions for the improvement of learning analytics within the evaluation planning and learning processes. And in particular learning analytics can support the enabling of new ways for the impact assessment of learning opportunities within a general evaluation framework that has to be discovered and discussed.

4. Conclusions

This paper presented how learning analytics can be included within a general evaluation framework and already be defined at the beginning of the learning design. For this specific purpose all open questions and issues concerning privacy and data protection that arise from a broad application of learning analytics were excluded: Such questions were outside of the scope of this paper, but must be addressed and are very important for a successful introduction of learning analytics. The generic Evaluation Framework for Impact Assessment (EFI) was introduced as an example of a general evaluation framework and could identify which part of it is covered by learning analytics as well as how learning analytics can contribute to the impact assessment. Based on the reference process model from the international quality standard ISO/IEC 19796-1, it was demonstrated that learning analytics can already be addressed during the start of the design phase for a learning opportunity. An extension of the Learning Design specification can contribute to the introduction and support of learning analytics. Future research should focus on the open question as to how the introduction of learning analytics can be harmonized in different systems and for different target groups and organizations by using standardized phases and processes such as the IDEAL reference framework. This would lead to comparable and hopefully interoperable learning analytics systems and data for the analysis and benchmarking across different systems, target groups and organizations.
References


Learning Analytics Data Items on Digital Textbooks

Yasuhisa Tamura*
*Sophia University, Japan
*ytamura@sophia.ac.jp

Abstract: This paper proposes a set of data items to be collected in Digital Textbooks working on desktop/laptop/tablet PCs. Based on conventional LMS-based learning activity analytics, various types of data were proposed to use. In addition, modern tablet PC-based learning has advantage to collect more detailed learner’s data with use of equipped sensors and material manipulation logging. This proposal is under discussion in IDPF EDUPUB community, which aims to specify ePub3-based Digital Textbook format and functions.

Keywords: Learning analytics, e-Textbooks, Learners’ behavior, Analysis, Sensors, EDUPUB

1. Introduction

This article focuses on combination of two modern issues: Learning Analytics and Digital Textbooks. Some background information of these issues are introduced below.

Learning analytics (LA) has become a major area in learning science and learning technology research. From the end of 1990s, LMS (Learning Management System) based learning environments have emerged. Since then, many types of learning activities logs have been collected in these LMSs and analyzed. These data come from instruction-based activities, e.g. class participations, material views, and answers to quizzes. Also they include active learning-based ones, e.g. enrollments, utterances, interim and final products of group activities.

For LA researches, there are a series of International Conferences on Learning Analytics and Knowledge. These proceedings are available: Long et al. (2011), Dawson et al. (2012), Suthers et al. (2013) and Pistilli et al. (2014). As a general survey, Shum (2012) classifies 5 types of LA activities: (1) analysis dashboard of LMS or VLE, (2) predictive analysis, (3) adaptive learning analytics, (4) social network analysis, and (5) discourse analysis. Especially for active learning and collaborative learning, Shum and Ferguson (2012) shows some LA goal and future issues of these activities. Up to date discussion and information are available on Google Groups on Learning Analytics (2014).

As described below, the author intends to establish a basic and standard collection of data items to use LA activities with use of Digital Textbooks. This collection should include data items that are utilized in previous researches. In order to clarify these data items, the author investigated some of previous published papers and listed up the used data items. The summary is shown in Table 1 and Table 2. Table 1 shows 17 papers to focus on classroom and individual learning. Also, Table 2 shows 13 papers to focus on collaborative and active learning. These data items are referred in the proposal in Section 2.

Digital textbooks, also known as e-textbooks, are now investigated and planned to implement at several countries all over the world. KERIS (2014) in Korea started investigation and experiment in 2008, and lead to finish implementation throughout the country until the end of 2015. Also China, Singapore, Philippines, India and other Asian countries are proceeding investigation and experimental introduction. In Europe, England, France, Germany, Spain and other countries are under investigation and experiment. In United States of America, some states including California, Washington and Utah are planning to deliver open textbooks or complementary devices.
Table 1. Data items and Objectives of Learning Analytics Researches
(classroom and individual learning).

<table>
<thead>
<tr>
<th>Reference</th>
<th>Data Items</th>
<th>Goal of Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arnold and Pistilli (2012)</td>
<td>Posting of a traffic signal indicator on a student's LMS home page, E-mail messages or reminders, Text messages, Referral to academic advisor or academic resource center, Face to face meetings with the instructor</td>
<td>Relationship between items and achievement</td>
</tr>
<tr>
<td>Barber and Sharkey (2012)</td>
<td>Prior credits earned, Discussion post count/week, Late assignments, Orientation participation, Count of messages to instructor, Inactive time since last course</td>
<td>Prediction of class achievement</td>
</tr>
<tr>
<td>Clow (2013)</td>
<td>Visit, Registration, and contribution ratio of MOOCs</td>
<td>Drop rate analysis of MOOCs learners</td>
</tr>
<tr>
<td>Graf et al. (2011)</td>
<td>Templates, patterns, learning object, database connections of materials</td>
<td>Judgment of material difficulty</td>
</tr>
<tr>
<td>Holman et al. (2013)</td>
<td>Grade, Class standing, and badges of quizzes</td>
<td>Self prediction of achievement</td>
</tr>
<tr>
<td>Kizilcec et al. (2013)</td>
<td>Visiting, Enrollment, and assessment numbers in MOOCs courses</td>
<td>Number transition of MOOCs learners</td>
</tr>
<tr>
<td>Lonn et al. (2012)</td>
<td>Grade information every few weeks</td>
<td>Assistance necessity from mentors</td>
</tr>
<tr>
<td>Martin et al. (2013)</td>
<td>Answers of each sub-quiz</td>
<td>Visualization of learning process</td>
</tr>
<tr>
<td>Monroy et al. (2013)</td>
<td>Teacher’s usage of teaching unit parts (overview, essentials, engage, explore, explain, evaluate, intervention, acceleration)</td>
<td>Heat map of unit parts usage</td>
</tr>
<tr>
<td>Niemann et al. (2012)</td>
<td>Learning object usage in a web portal</td>
<td>Similarity of learning objects</td>
</tr>
<tr>
<td>Pardos et al. (2013)</td>
<td>Quizzes and scaffolding help</td>
<td>Relationship between Scaffolding help and achievement</td>
</tr>
<tr>
<td>Raca and Dillenbourg (2013)</td>
<td>Video captured actions of learners</td>
<td>Learner behavior during classrooms</td>
</tr>
<tr>
<td>Santos et al. (2012)</td>
<td>Date and time range of learners</td>
<td>Visualization of learning status</td>
</tr>
<tr>
<td>Sao Pedro et al. (2012)</td>
<td>Quiz answers</td>
<td>Transition of problem solving skills</td>
</tr>
<tr>
<td>Tempelaar et al. (2013)</td>
<td>Achievements in various learning areas</td>
<td>Skill analysis (Self-belief, learning focus, planning, management, persistence)</td>
</tr>
<tr>
<td>Verbert and Duval (2011)</td>
<td>Dataset and functions of recommender system</td>
<td>Comparison of Recommender systems</td>
</tr>
<tr>
<td>Wolff and Zdrahal (2013)</td>
<td>Precision and recall of learning units</td>
<td>Comparison of TMA (Tutor-marked assessment) and VLE (Virtual learning environment)</td>
</tr>
<tr>
<td>Reference</td>
<td>Data Items</td>
<td>Goal of Analysis</td>
</tr>
<tr>
<td>---------------------------</td>
<td>----------------------------------------------------------------------------</td>
<td>----------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Ahn (2013)</td>
<td>Emails received, Emails sent, Friends, Friend Lists, Links, Member pages, Networks, Notes, Photos, Status messages, Videos, Wall posts</td>
<td>Factor analysis of media literacy (Negotiation, Networking, Judgment, Play, Multitasking, Appropriation, Transmedia navigation)</td>
</tr>
<tr>
<td>Cambridge and Perez-Lopez (2012)</td>
<td>Discussion post, blog, their narratives,</td>
<td>Analysis of discourse style to activate learner groups</td>
</tr>
<tr>
<td>Camilleri et al. (2013)</td>
<td>Pleases and numbers of utterances in virtual space</td>
<td>Behavior analysis</td>
</tr>
<tr>
<td>Cobo et al. (2012)</td>
<td>Reading and writing activities during online discussions</td>
<td>Clustering of learners</td>
</tr>
<tr>
<td>Ferguson and Shum (2011)</td>
<td>Keywords in text chat</td>
<td>Chat type (evaluation, explanation, reasoning, justification, perspective)</td>
</tr>
<tr>
<td>Koulocheri and Xenos (2013)</td>
<td>Bookmarks, blog posts, topics and files uploaded, bookmarks, comments on bookmarks/blog posts/topics/files in group</td>
<td>Visualization of member relationships</td>
</tr>
<tr>
<td>De Liddo et al. (2011)</td>
<td>Response type of utterances (respond, about, example, solution, support)</td>
<td>Relationship analysis of learners</td>
</tr>
<tr>
<td>Schneider et al. (2013)</td>
<td>Eye-tracking data</td>
<td>Estimation of collaborative learning skills</td>
</tr>
<tr>
<td>Schreurs et al. (2013)</td>
<td>Person, type of tie, topic</td>
<td>Visualization of learner relationship network</td>
</tr>
<tr>
<td>Shum and Crick (2012)</td>
<td>Quiz achievement and various activities</td>
<td>Relationship between individual learning achievement and meta-skills</td>
</tr>
<tr>
<td>Siadaty et al. (2012)</td>
<td>Vocabulary in shared Wiki and bookmark</td>
<td>Collaborative skills analysis of corporate learners</td>
</tr>
<tr>
<td>Tempelaar et al. (2013)</td>
<td>Achievements in various learning areas</td>
<td>Analysis of necessary skills (Self-belief, learning focus, planning, management, persistence)</td>
</tr>
</tbody>
</table>
In Japan, MEXT (Ministry of Education, Culture, Sports, Science and Technology) (2011) published a roadmap called “The Vision for ICT in Education”, which were planning to introduce digital textbooks countrywide until 2020. Also, an experimental project was deployed from 2011 to 2013. It was a joint project between MEXT and MIC (Ministry of Internal Affairs and Communication) to introduce ICT and digital learning materials to selected 20 schools. Final report of this project (in Japanese) is available through MEXT (2014). At the same time, MEXT and MIC started experimental development projects of Digital Textbooks in 2013. In these projects, MEXT is focusing ePub3, while MIC is HTML5. These projects will continue in 2014.

On the other hand, various standardization organizations and communities are trying to specify standard file formats and specifications for Digital Textbooks. These projects and their timelines are shown in Figure 1. IEEE (2014) initiated Actionable Data Book Project in 2011, and published some research papers. Also, CEN (European Committee for Standardization) (2014) and IMS Global Learning Consortium (2014) began eTernity Project and ICE Project in 2012 and 2013, respectively. Among them, ISO/IEC JTC1/SC36 (2014), a subcommittee of ISO dedicated to e-learning technical specifications, started e-Textbook Project in September 2012 meeting at Busan, Korea. It is investigating related standardization activities, issued a set of questionnaires of Digital Textbooks to standardization communities, and arranged future issues in a document in 2014 meeting.

The latest and the most active one is called EDUPUB project. It is lead by IDPF (International Digital Publishing Forum) (2014), which specified ePub3 format for Digital Books. The first workshop of EDUPUB was held in October 2013 at Boston, USA, while the second in February 2014 at Salt Lake City, USA. The third workshop was held in June 2014 at Oslo, Norway, and the fourth is scheduled in September 2014 at Tokyo, Japan. Through these workshops, these outlines below are discussed.

- Core file format is ePub3.
- In order to add textbooks specific structural semantics, Pearson and Benesse staffs proposed their textbook and material descriptive tags. It is under online discussion.
- In order to attach learner note into the textbooks, “Open Annotation in ePub” specification is under discussion.
- For quiz data format, IMS QTI (2014) (Question and Test Interoperability) specification is a major candidate.
- For calling scheme of outer applications or resources, IMS LTI (2014) (Learning Tools Interoperability) specification is a major candidate.
- Textbook specific metadata items are under discussion.

The author is a member of MEXT Digital Textbook project, ISO/IEC JTC1/SC36 e-Textbook project, and IDPF EDUPUB Project. In the EDUPUB project, there was a proposal to specify a set of data items to be collected with use of LA. For this proposal, the author started to survey conventional research papers in order to specify commonly used data items, and also proposed a new set of items which are able to collect with use of tablet PCs. Section 2 shows this proposal.

2. Data Items Acquired with use of Digital Textbooks

2.1 Characteristics of Tablet PCs
There are many types of PCs utilized in classroom and individual learning. Both desktop PCs and laptop PCs have been common. In addition, tablet PCs have become popular in these years. Apple launched a first iPad in 2010. Also, Google and China/Korean hardware companies began to launch Android based tablet PCs in 2010. Nowadays, worldwide shipments of desktop / laptop PCs and tablet PCs are almost equal in 2014. In 2015, shipments of tablet PCs will be 20% more than desktop / laptop (Gartner 2014).

Even traditional laptop PCs are able to connect to computer network, download information from a certain server, upload it to a server, or communicate each other with use of e-mail and SNS. Also they have some sensors: brightness sensor, camera, and microphone. With use of these functions, they are able to generate data to be used in LA:

- Enroll to a class in LMS
- Access materials in LMS
- Upload quiz answer / assignment / reaction
- Show hint / advise
- Send, receive and read messages from / to instructors
- Enroll to a group in LMS
- Send / receive text / audio / video messages from / to instructors / another learner
- Access to shared whiteboard / file
- Timestamp of these activities

In conventional way, all of these information are collected in LMS. With use of additional functions attached with learners’ Web browsers, some information can be collected in client (=learner) PCs, but this approach is not common.

However, modern tablet PCs equip many other types of sensors: screen touch sensor, GPS, digital compass, gyroscope, acceleration sensor, etc. With use of these sensors, a tablet PC is able to collect various information about learning activities and their environment. For example:

- View / flip one’s textbook, reference or dictionary
- Insert highlights or underlines in one’s textbook, reference or dictionary
- Write notes or annotations on one’s textbook, reference or dictionary
- Refer reference or dictionary by specifying a certain part of textbook
- In addition to timestamps, places of these activities
- Environmental voice and noise of these activities
- Learner’s face, expression, and visual environment of these activities

Some of these data are collected with use of equipped sensors directly, others should be analyzed by Digital Textbook viewer software or related application software. Also, the data collection implies privacy violation. This issue will be discussed in the later section.

2.2 Proposed Data Items

Based on investigation and consideration stated above, the author proposes data items below to be collected with use of Digital Textbooks. Figure 1 shows a framework. One atomic data includes “who” (Subject), “when” (Date & Time), “where” (Geographic point location, optional), and “what”. A detail of lower right side table of Figure 2 is shown in Table 3.

Data items in Table 3 consist of two categories: (1) commonly used in conventional LA researches, shown in Table 1 and Table 2, and (2) assumed to be collected on Tablet PCs, Digital Textbook viewer and related software mentioned in Section 2.1.

This proposal is now disclosed to EDUPUB community, and weekly discussion is ongoing.
Table 3. Detail of Proposed Data Items: Verbs and Objects.

<table>
<thead>
<tr>
<th>Category</th>
<th>Verb (action)/ Object (target)</th>
<th>attend/quit</th>
<th>flip/view</th>
<th>add</th>
<th>modify</th>
<th>delete</th>
<th>answer</th>
<th>send</th>
<th>receive</th>
</tr>
</thead>
<tbody>
<tr>
<td>Classical/individual Learning</td>
<td>Class</td>
<td>o</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Page of e-textbook or reference</td>
<td>o</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Highlight / underline</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Note (annotation)</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Link</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Quiz</td>
<td>o</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Assignment</td>
<td>o</td>
<td></td>
<td></td>
<td>o</td>
<td>o</td>
<td>o</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Feedback</td>
<td>o</td>
<td></td>
<td></td>
<td></td>
<td>o</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Message</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>o</td>
<td>o</td>
</tr>
<tr>
<td>Collaborative and active learning</td>
<td>Group</td>
<td>o</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Shared whiteboard</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Shared file</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td></td>
<td></td>
<td>o</td>
<td>o</td>
<td></td>
</tr>
</tbody>
</table>

3. Discussion and Conclusion

During development process of the draft proposal above, there were some discussions whether tablet PC based fine-grained data should be included, for example:

- Face expression of a learner,
- Attitude of a learner,
- Voice of a learner and environmental sound,
- Acceleration data,
- Digital compass data, and
- Gyroscope data.
Also, it is emerging to utilize so called “wearable devices” to collect biological and environmental data, for example:

- Temperature of learner’s body and environment,
- Humidity of environment,
- Body sweat of a learner,
- Heart rate of a learner,
- Blood pressure of a learner,
- Eye-tracking data of a learner, and
- Brain waves of a learner.

Currently, it is not clear that these data are useful to identify learner’s status or not, at least from preceding research results. Therefore, the author thinks that the data listed above are still early to include as “standard” data items for learning analytics. It’s why the listed data is omitted from Table 3. However, as far as the author knows, the proposal appeared in Figure 2 and Table 3 is the first appearance for “standard” data items for learning analytics activities. This “standard” means that major stakeholders support to adapt them as useful ones. There are many goals in LA, shown as samples in Table 1 and Table 2. This proposed data items cover these, and similar LA goals.

A major discussion point for the proposed data items is a risk of privacy violation of learners. This proposal includes geographical data and timestamp. So, an analyst or an instructor is able to grasp when and where a learner is. Also, a tablet PC is able to collect visual and audio data during learning activities. It might clarify a scene and accompanying friends during learning. Currently it is not clear what data violates learner’s privacy and doesn’t. We should clarify a threshold of private data, and make broad consensus. From this viewpoint, Table 3 does not include visual and audio data not intentionally recorded by learners.

One of the future issues is comprehensiveness of the proposal. Currently there is no major argument for the proposal. However, there are many other existing researches of LA. They should be investigated in order to guarantee comprehensive of this proposal. Also, IMS proposes Caliper specification. It also specifies a data set for LA. The document is not opened, but needs investigation.

The other one is to verify usefulness of these data items for LA activities. The proposed data items are listed from viewpoint of technical feasibility with use of tablet PCs and Digital Textbook software. However, it is not clear what characteristics can be analyzed with use of these items. Some them are already clear based on the conventional researches in Table 1 and Table 2. However, especially Digital Textbook specific items should be verified in this usefulness.

To conclude, the author proposes a set of data items to be collected with use of tablet PCs and Digital Textbook viewer software, shown in Figure 1 and Table 3. They are far more detailed than conventional LMS based data collection. However, it should be enhanced and brushed up in order to guarantee comprehensiveness and learner’s privacy.

Acknowledgements

This work was supported by JSPS KAKENHI Grant Number 26282059. The author would like to appreciate Markus Gylling and all EDUPUB project members for dedicated discussion and continuous encouragement for this work.

References


Learning Analytics: An Enabler for Dropout Prediction

Shu-Fen TSENG\textsuperscript{ad}, Chih-Yueh CHOU\textsuperscript{bd}, Zhi-Hong CHEN\textsuperscript{cd}, Po-Yao CHAO\textsuperscript{cd}
\textsuperscript{a} Department of Information Management, Yuan Ze University, Taiwan
\textsuperscript{b} Department of Information Science & Engineering, Yuan Ze University, Taiwan
\textsuperscript{c} Department of Information Communication, Yuan Ze University, Taiwan
\textsuperscript{d} Innovation Center for Big Data and Digital Convergence, Yuan Ze University, Taiwan
*gssftseng@saturn.yzu.edu.tw

Abstract: A key application of learning analytics is predicting students’ learning performances and risks of dropping out. Heterogeneous data were collected from selected school to yield a model for predicting student’s dropout. Results from this exploratory study conclude dropout prediction by learning analytics may provide more precise information on identifying at-risk students and factors causing them to be at risk.

Keywords: Learning analytics, dropout, predictive model

1. Introduction

It is gradually recognized the effective use of data and learning analytics are both critical components of digital learning strategies to personalize learning for needed students to increase student retention and achievement in schools (Bienkowski, et al., 2012). Analytics applied to education data can help schools and school systems to better understand how students learn and succeed. Significant improvements in technology tools and resources, and the focus on meeting the needs of individual students through personalized and digital learning have together provided an emerging context in which education systems have the opportunity to advance the values of learning analytics to truly inform teaching and learning.

A key application of learning analytics is monitoring and predicting students’ learning performance and addressing potential issues early so that interventions can be provided to identify students who are at risk of failing a course of program of study (Bienkowski, et al., 2012). In particular, one of the key successes of learning analytics is predicting which students are at risk of dropping out. Tobin and Sugai (1999) demonstrated that using data on disciplinary referrals during middle school can predict those who are likely to drop out of high school. Their research supports the use of school records of discipline referrals as a screening device. The work by Bowers (2010) shows that by analyzing student grade history from k-8 or k-12, it can correctly identify over 80% of dropout students.

One of noteworthy cases is that IBM worked with Mobile County Public Schools in Alabama to apply analytics to education data to help school system identify which students were at risk of dropping out. Rapid access to information and analytics means school administrators and principals can make more informed decisions and take appropriate and timely action to develop an individualized response to each student’s problems and monitor their progress. Timely intervention based on real-time information is helping Mobile County to keep students on the right path and to lower dropout rates (Centre for Information Policy Leadership, 2013; IBM, 2009).

It has long been recognized by school authorities in Taiwan that dropout students mostly come from disadvantage family (e.g. single parent, minority), who lack of self-discipline, are not interested in school work, and easily befriended juvenile delinquents. There were a total of 5,379 dropouts among k1-k12 students in the year of 2012 in Taiwan. Personal profiling outlined mostly of them dropped out during their middle school periods (89%). Among these dropouts, 58% of them came from single parent household. Major causes of their dropout were categorized into individual reasons (47%), followed by family (24%) and social (17%) problems (MOE, 2012).
School administrators and educators can only analyze these students when they had already dropped out of school. In contrast of taking remedial actions, the newly development of learning analytics could help predict in advance and provide the tools to calculate a student’s risk of dropping out of school. In this exploratory study, we collect students’ data from heterogeneous sources (ie. grade history, disciplinary referral, class attendance) and employ logistic regression model to identify risk factors of dropping out in one selected school. Then, we compare the findings to profiles of dropout that outlined by conventional method. We further discuss the potential benefits of learning analytics on dropout prediction.

2. Method

Taoyuan county ranks the second highest dropout rate in Taiwan. In this study, we collected data from one middle school in Taoyuan county. The school selected was among one of the highest dropout rate schools. Heterogeneous sources of administrative, academic, and disciplinary data of every student in this school were collected and organized into a database. Descriptive and explanatory statistics were used to determine the relationship between a dependent variable (whether a student was dropped out in current term) and independent variables (e.g. students’ family background, academic grade, class absence, and disciplinary referral records in previous term). In particular, logistic regression analysis was employed to create a model for predicting whether a student was at risk of leaving school.

3. Findings

In the year of 2013, there were a total of 712 seventh and eighth grade students in this middle school. Among them, 11 students dropped out of school. Five of them were males and four out of 11 students were minorities. Correlation coefficients showed that dropout was positively related to number of class absence ($r=.363$), number of disciplinary referral ($r=.253$) in previous term, and it was negatively related to academic grade ($r=-.06$) of previous term. Despite the limited number of dropout cases, logistic regression model (table 1) still showed prediction power of disciplinary referral (odd ratio=1.136) and academic grade (odd ratio=.932) of previous term on a students’ risk of dropping out of school. Number of class absence in previous term and gender showed insignificant impacts on predicting dropout.

| Constant | .332 | 1.394 |
| Male | -.409 | .664 |
| Class Absence (previous term) | .013 | 1.013 |
| Disciplinary referral (previous term) | .127 | 1.136* | Increase risk by 13.6% |
| Academic grade (previous term) | -.07 | .932** | Decrease risk by 6.8% |

4. Discussion

The conventional way of outlining high risk student of dropouts is to analyze demographic and family characteristics of those who had already dropped out. In the year of 2013, among those dropout students in Taoyuan county, 47% of dropouts were from single parent household, 24% of them were minorities, and 6% of them were raised by grandparents. Based on this conventional method, individual and family reasons were ranked the highest causes of student to drop out of school.
By employing method of learning analytic, we collected and organized data from personal and family composition, academic performance and disciplinary records; our finding suggests that we can predict a student’s risk of dropping out of school mainly based on his/her school related activity records of previous term. In particular, high number of disciplinary referral and poor academic grade were major factors that help to predict student’s risk of dropping out of school. The results suggested instead of targeting individual reasons and family background, student’s activities in school can be effectively used to predict a student’s risk of dropping out of school. With more longitudinal data of students are collected and more middle schools join the project, we could build a better fit model for dropout prediction and calculate risk of dropout for students. This exploratory study concludes that dropout prediction from learning analytics may provide more precise information on identifying at-risk students and factors causing them to be at risk.

References

Ministry of Education (2012). Dropout statistic in Taiwan. Executive Yuan, Taiwan.
Motivation and Engagement in MOOC – Teachers’ Perspective

Li Fern TAN*, Kai Song GOHb & Emile SABASTIANb
aOffice of Learning Technologies, Temasek Polytechnic, Singapore
bSchool of Informatics & IT, Temasek Polytechnic, Singapore
*lifern@tp.edu.sg

Abstract: In recent years, the growth of online educational programs has been stimulated by the advancement of the Internet and learning technologies which have transformed the educational landscape. Massive Open Online Courses (MOOCs) have gained much popularity over the past few years and have changed the way people learn. In this qualitative study, we interviewed 14 academic staff at a polytechnic in Singapore and examined the key factors that motivated them to sign up for learning via MOOCs as well as factors that affected their choice of MOOC subjects. Our participants consisted of two distinct groups of academic staff, those who completed the MOOCs they signed up for and those who did not complete. We discovered that the two groups of participants were motivated differently when signing up for MOOCs. We also investigated the factors that led to successful completion of MOOCs by one group of academic staff and the challenges faced by the other group that cause their incompletion. Finally, we asked academic staff for their recommendations on how they think MOOCs could be made more engaging and adaptive to learners with different learning needs and styles. Their recommendations on how MOOCs should be administered, delivered and assessed are presented in this study.

Keywords: MOOC, motivation, engagement, adaptive

1. Introduction

Massive Open Online Courses (MOOCs) cast education in a new paradigm, leveraging on the pervasiveness of Web access and the potential of social learning. By making education available to everyone with access to the web, MOOCs break the boundaries of economic access, time and geographical location faced by traditional education. What was only available to a privileged group of learners is now available to tens of thousands of students. A MOOC generally carries no charges, no prerequisites, no predefined expectations for participation and no formal accreditation. It aims to reach out to anyone who is interested to learn something new. Today, in addition to taking some MOOCs without charge, participants may pay a fee for courses that lead to a certificate (Waldrop, 2013).

Online learning through MOOC is one of the emerging technologies for learning in education. According to Wikipedia (2014) emerging technologies learning is defined as “technical innovations that brings out new territory in some significant way”. MOOC has evolved as an emerging technology into new pedagogy that will have an impact on teaching and learning. NMC Horizon Report (2013) stated that higher education will see widespread adoption of MOOC and tablet computing.

The pedagogy that MOOCs employ is different from traditional online learning. MOOCs can be delivered synchronously on a predefined schedule allowing learners to do his or her lessons without geographical boundaries and at his or her convenience. Instead of making a 45-60 minute recorded lecture online for learning, short lecture modules, each lasting 12-15 minutes are used. The shorter duration provides learners with convenient blocks to complete the learning that could be easily fitted in their schedule. Assessments and grading are automated within the platform. MOOCs are used for continuing education objectives by learners who wish to supplement their previous education with skills-enhancement or personal challenge purposes.

MOOCs began in 2011 when Stanford professor, Sebastian Thrun and Director of Research at Google, Peter Norvig taught “Introduction to Artificial Intelligence” online with 160,000 students enrolled and 20,000 completing it. Professor Thrun later resigned and founded Udacity. Thrun’s
colleagues at Stanford, Daphne Koller and Andrew Ng established their own platform, Coursera, in February 2012. Harvard’s and MIT’s EdX started in May 2013. Other MOOCs soon followed in the wake of these MOOCs (Nanfito, 2014). Today, institutions are offering a variety of courses on MOOC platforms such as Coursera, Canvas Network, Udacity and EduKart. There are other MOOC-like online platforms like Fathom, Sunoikisis and Connexions.

Enrolments in online education are increasing substantially although retention and completion rates remain low in the face of declining enrollment in higher education (Allen & Seaman, 2013). The 2013 Babson report by Allen and Seaman (2013) shows that more than six million students in public, private and for-profit educational institutes in the United States took at least one online course in the fall 2011 term. Other key findings from the report include low completion rates is barrier for the growth of online learning and 88.8 percent of academic leaders surveyed believe that student lack of discipline in online courses is an obstacle to growth. Cognitive, psychological and emotional connections to feel, think and behave are required for the online environment (Lehman & Conceicao, 2010). Concerns for motivation and factors for engaging learners have to be taken into consideration to help students stay motivated online. It is understandable that a few common reasons for student dropout are related to feelings of isolation, technology disruption, lack of support from faculty, lack of clarity in instructional direction, lack of social interaction and so on.

Learners in the twenty-first century have been Web consumers for much of their lives, and are now demanding online instruction that supports participation and interaction. They want learning experiences that are social and that will connect them with their peers. West & West, 2009, p.2

It is also important to understand learners’ characteristics, their learning behaviors and skills in the 21st century. The ECAR Study of Undergraduate Students and Information Technology, 2012 (EDUCAUSE, 2012) indicated almost 9 in 10 students own laptops, more than 60% of the students own smartphones and 15% of the students possess a tablet in United States. In a survey conducted by Infocomm Development Authority of Singapore (2014) in the year 2012, 99% of the individuals in the age group between 15 - 24years old who have used a computer and used the internet for the past 12 months. This reflects the behaviour of the generation of learners born in this digital era. Changes in student behavior due to technology usage bring new demands for learning and teaching. Students in the 21st century are IT savvy and are comfortable with technology. The use of Internet has become a norm and may be a way of life for students. They demand greater autonomy of their own learning and the addition of technologies has met their learning needs and preferences (Prensky, 2005). The infusion of information and communication technologies in teaching and learning has open up a wide range of opportunities for creating new kinds of learning activities and experiences. Technology is no longer the problem, but what to do with them to succeed in the new learning environment in this digital era (Carr, 2011). The advancement and adoption of technology for teaching had also transformed the role of the teacher. Kearsley (2000) wrote that the role of the instructor in online classes is to ensure high degree of interactivity and participation through careful design of learning activities which result in engagement with the subject matter and with the students. We felt that there is a need to explore ways to motivate and engage students to help them succeed in the online classroom.

2. Purpose of the Study

Motivation is one of the critical success factors leading to course completion in MOOCs. Motivated learners are more likely to engage in learning activities, participate in online discussion and ultimately, succeed in the course. Thus it is important to understand what motivates online learners, especially academic staff. In this study, key factors that influenced academic staff in their motivation and engagement in MOOCs and how these factors were embedded in the design of learning elements in MOOCs would be investigated. Furthermore, the contributing factors that have motivated academic staff to complete or have hindered their learning progress in MOOCs would be discussed as well. With these in mind, the research questions explored in this study are as follows:

1. What were the contributing factors that led the academic staff to sign up for MOOCs?

2. What were the contributing factors that led to successful completion of MOOCs among the academic staff?
3. What were the contributing factors that caused academic staff to drop out of MOOCs they had signed up for?
4. How had the instructors of the MOOC adapted to the different learning needs of participants?

3. Methodology

3.1 Method

Semi-structured interviews were used to collect data for this research. This qualitative inquiry is well suited for educational research as it enables deep exploration. Interviewees have the freedom to share their experiences and the interviewer retains control of the interview at the same time (Drever, 1995). It also provides the interviewer the freedom to explore general views or opinions in detail (Robson, 2002). Prior to the interview, pilot interviews were conducted to ensure that the set of questions used was effective in fulfilling the purpose of this study.

3.2 Participants and Settings

This study took place at a polytechnic, post-secondary institute, in Singapore. This is one of the five polytechnics in Singapore. There are six schools of study in this polytechnic each offering several diplomas in their specific domain of study. Purposive sampling was used to choose the sample consisting of participants who were appropriate for the study to provide rich information for researcher to develop a detailed analysis on the central phenomenon under study. The research participants were categorized into two groups; those that obtained statements of completion for the MOOCs they signed up for, we call this group of participants the “Completed” group; and those who went through at least 10 hours of the MOOC they signed up for but did not obtain statements of completion, we term this group of participants the “Attended” group. All participants were academic staff from the same school at the polytechnic, each teaching different subjects offered by the school.

Sampling was done according to the matrix as shown in Figure 1. We selected at least three participants from each category to ensure that each category was represented. Each category consists of a fair mix of appointment and non-appointment holders. Appointment holders refer to academic staff who hold an academic appointment like course manager, course coordinator or section head for the course. Non-appointment holders refer to academic staff whose main focus is on teaching. Both appointment and non-appointment holders are heavily involved in academic related matters, from course planning, course design to delivery, we believe that their inputs would be beneficial to this study. In all, 14 staff members were interviewed for our research.

<table>
<thead>
<tr>
<th>Staff Type</th>
<th>Completed Group (Obtained Statement of Completion of any MOOC)</th>
<th>Attended Group (Completed more than 10 hours of any MOOC)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Appointment Holder (Managers/Coordinators/Section Heads)</td>
<td>At least 3 participants</td>
<td>At least 3 participants</td>
</tr>
<tr>
<td>Non-appointment Holder</td>
<td>At least 3 participants</td>
<td>At least 3 participants</td>
</tr>
</tbody>
</table>

Figure 1. Quota sampling criteria of research participants.

4. Findings

Our participants enrolled in MOOC for a variety of reasons, some have more initial interest in upgrading themselves than others. A key principle to the framework of self-determination theory (Ryan & Deci, 2000) is that individuals enjoy activities when they believe they have autonomy over some
aspects of them. Individuals who are self-determined perceive they have ability to make choices over their actions and have been shown to have augmented conceptual learning, positive attitude towards challenging tasks and increased motivation to attend lessons which resulted better performance (Filak & Sheldon, 2008). When it came to the motivation for signing up for learning via MOOC, there were noticeable differences between the “Completed” group and the “Attended” group.

Many of the academic staff who completed their MOOCs cited professional curiosity as one of the factors that prompted them to sign up for MOOCs. They were keen to find out how a course could be delivered in a fully online mode, to a large number of participants; operational details such as how lessons were organized and delivered and how assessments were administered and graded, how queries could be responded to were also on their lists. In short, they showed enthusiasm to find out how MOOCs worked.

While most academic staff from the “Attended” group signed up for MOOCs with the initial intention of completing it, all eventually did not complete. Only 1 interviewee from this group had foreseen that he would be unable to complete the course when he signed up for the MOOC, but did so anyway with the intention to explore the subject matter. The group of academic staff who did not complete the course mainly took on learning via MOOC because they believe this mode of learning suited their own work schedule. They also saw MOOCs as a rich source of up-to-date materials for subject area of their interest.

Some common motivational factors among the two groups of academic staff include fulfilling work-related goals, taking MOOCs as a personal challenge and as a form of self-enrichment. Only 1 of the interviewees mentioned that he signed up for learning via MOOC because it was free. This suggests to us that the flexibility of choice and convenience of having learning delivered via the Internet outweighs the draw that MOOC was delivered at no costs. Studies have shown that students’ motivation is affected by their perception of the usefulness of what they would have been taught (Tabachnick et al., 2008). It was also found that students with long term goals or involved with long term projects who are able to see the significance or bearing in their learning with their future are more motivated as compared to those having short term goals.

The Venn diagram below summarizes the factors that prompted academic staff to sign up for learning via MOOC. There were no noticeable difference between appointment holders and non-appointment holders.

![Venn Diagram](image)

**Figure 2.** Factors that prompted staff to sign up for learning via MOOC.

We asked our interviewees what motivated them to sign up for the particular MOOCs they did. Even though the topic areas of the MOOCs they took ranged quite widely, the reasons why they chose what they did were fairly consistent when we fed the captured interview transcript into a word cloud generator.

Word cloud shown in Figure 3 was used to identify the prominent terms that were gathered from the interviews. Featured prominently were terms like “reputable”, “university”, “interest”, “subject matter” and “relevance” which suggested that our participants preferred MOOCs offered by reputable institutions on subject matter that were related to their area of work or area of interest.
For the group of academic staff who completed their MOOCs, we asked them what helped kept them going, we classify these into intrinsic and extrinsic factors. A summary of the factors, both intrinsic and extrinsic, that helped academic staff we interviewed complete the MOOCs they took can be found in Figure 4.

<table>
<thead>
<tr>
<th>Intrinsic Motivation Factors</th>
<th>Extrinsic Motivation Factors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sense of achievement upon completion.</td>
<td>Certificate or badge received upon completion.</td>
</tr>
<tr>
<td>Relevance and applicability of subject matter to own work area.</td>
<td>Regular email reminders of upcoming tasks.</td>
</tr>
<tr>
<td>Interest in subject matter.</td>
<td>Availability of course calendar showing important dates.</td>
</tr>
<tr>
<td>Sense of being challenged by the MOOC.</td>
<td>Organization and design of materials.</td>
</tr>
<tr>
<td>Curiosity of how the MOOC will end.</td>
<td>Sequencing of topics.</td>
</tr>
<tr>
<td>Peer and social support.</td>
<td>Right pitching of demands of assessments.</td>
</tr>
</tbody>
</table>

Figure 4. Factors that helped academic staff to complete MOOCs they signed up.

The most cited intrinsic factors were the anticipated sense of achievement upon completion of the MOOC and the relevance and applicability of subject matter to own work area.

The most cited extrinsic factors that kept them going were the alignment of the MOOCs’ schedules to their own work schedule as well as the course design elements such as the organization and design of materials, sequencing of topics, right pitching of demands of assessments. Course administrative tools such as regular email reminders and availability of course calendar also helped our interviewees plan their schedule ahead to accommodate important submissions and to manage workload for their MOOCs.

For the group that did not complete their MOOCs, we found out what were some challenges they faced. These are classified into 2 categories, personal and course-related as shown in Figure 5.

<table>
<thead>
<tr>
<th>Personal challenges</th>
<th>Course-related challenges</th>
</tr>
</thead>
<tbody>
<tr>
<td>Work commitment got in the way.</td>
<td>Course had too many assessments.</td>
</tr>
<tr>
<td>MOOC assumed prior knowledge which staff did not have.</td>
<td>Demands of MOOC were misrepresented.</td>
</tr>
<tr>
<td>MOOC turned out to be uninteresting.</td>
<td>Coverage of MOOC was misrepresented.</td>
</tr>
<tr>
<td>Assessments were too complex.</td>
<td>MOOC assumed prior knowledge which staff did not have.</td>
</tr>
<tr>
<td>Could not understand instructor’s accent.</td>
<td></td>
</tr>
</tbody>
</table>

Figure 5. Challenges faced by academic staff.
Unsurprisingly, many of our interviewees who did not complete their MOOCs failed to do so because of work commitments, this was particularly apparent for the appointment holders among the interviewees. Recall that all our interviewees were academic staff, which meant that there was less flexibility in making time for MOOCs since they were bounded by their own teaching time-tables and the general academic calendar. Some interviewees also made conscious decisions to give up on MOOCs which they felt were misrepresented in terms of time commitment, assessment demands and/or topics covered.

Based on their experience of learning via MOOCs, both categories of our interviewees agreed that at present, MOOCs are not able to adapt to the different needs and learning styles of learners. We then asked our interviewees for suggestions on how they would design MOOCs such that different learning needs and styles could be catered for. Their recommendations were summarized in the next section.

5. Recommendations

In this section, some recommendations based on the academic staff experience as a learner of MOOCs are collated from the research findings.

5.1 Provide flexible start dates for MOOCs

“Most of our students today are older, are working and need more flexible schedules” (Palloff & Pratt, 2001, p.109). At present, MOOCs are either self-paced or institution-paced. In the former, students can choose to start the course at any time, there are often no deadlines for assessments and all materials are available to student once he begins the course. The drawback of this is that students would be at different stages of completion compared to his peers. The social learning aspect of MOOC would diminish. There will also be no reminders on upcoming tasks or deadlines. In the latter arrangement, institution offering the MOOC will decide when the course will start. The problem with this design is that students’ busy period may coincide with the submission deadlines of the MOOC. Our research participants recommends for MOOC providers to consider providing courses with flexible start dates but not on a self-paced mode. This meant that students would have the flexibility to choose when they will start the MOOCs, once started the platform will work out a personalized calendar based on the MOOC’s original design. This would give students greater autonomy in deciding when they would embark on their learning. This would not only increase learner’s motivation through autonomy-supportive practice (Reeve & Jang, 2006) but also mitigating the problem of schedule clashes.

5.2 Provide flexible duration for MOOCs

An alternative to having flexible start dates for MOOCs is for MOOCs to have flexible durations. Presently, all institution-paced MOOCs also have institution determined duration. Our research participants recommend that MOOC providers allow different students to have different course duration for the same MOOC. The duration could be self-determined, based on learner’s assessment of the demands of the MOOC and their own aptitude; or it could be determined based on students’ performance for a diagnostic test, administered in the early weeks of the course. This provision of perceived control over the duration of the MOOC they are taking would be beneficial for students to stay motivated in their learning (Ryan & La Guardia, 1999).

5.3 Provide different track of study

Our research participants recommend that MOOC providers allow learners to determine their desired track of study with differing levels of difficulty. A learner who wishes to have a gentle introduction to the subject matter could opt for an introductory track while another who wishes for in-depth knowledge of the subject matter could opt for an advanced track of study. Learners have autonomy or sense of choice and feel controlled over their actions are more self-determined (Reeve et al., 2003). Some
aspects of the materials and assessments could be over-lapping; students of different tracks could interact and learn from each other through the existing collaboration platforms such as discussion forums.

5.4 Standardize instructional and presentation format

Many of our research participants had taken more than one MOOC. Feedback received indicated that different instructors organize their materials in vastly different manner. Our research participants recommend for MOOC instructors to adopt a standardized format by which materials are organized. Our participants believed that this consistency in structuring how course information and materials are presented would have a positive impact on their learning. Having a standardized and consistent instructional or presentation format will make it easier for learners to create a mental image of what to expect from the course and help them manage course workloads. We recognize that this would require extra organization effort by the instructor to rework and restructure instructional style but believe that this would be a worthwhile endeavor since an organized learning environment that provides relevant, consistent, practical and timely materials to meet learners’ needs, following the principle of easy to use and simple to use are important aspects to keep learners motivated.

5.5 Provide transcript for video lectures or audio lectures

Aside from recommending standardization of presentation and instructional format, our research participants also recommend that MOOC instructors consider providing a variety of learning materials. At present, materials for MOOCs mostly take the form of video lectures and lecture notes. Our research participants suggest for audio files and transcripts of video lectures to be made available as well. The former would be more suitable for people who prefer to learn on the go, using their mobile devices while the latter would cater to the group of learners who prefer to read rather than watch videos. Certainly, more effort would be needed to prepare the materials and it is more difficult to make any changes to the materials in future.

5.6 Create more opportunities for collaborative learning

The learning community is the vehicle through which learning occurs online. Members depend on each other to achieve the learning outcomes for the course. Palloff & Pratt, 2007, p.40

Instructors could consider incorporating more opportunities for collaborative learning in the course design. In the constructivist perspective, learning is being viewed as an active process whereby construction of knowledge takes place through social interactions and collaborative work with each other (Vygotsky, 1978). Students grasp their own understandings and construct knowledge through interactions based on what they already know and believe (Richardson, 2003). Moreover, students should be able to choose their collaborative learning partners. They were more motivated if they have the freedom to choose their working partners as compared to group assigned by the instructor (Ciani et al., 2008).

5.7 Provide intelligent progress tracking

Presently, most MOOCs do not track the progress of individual students. Progress bars are typically associated to the course schedule rather than students’ progress and email reminders are generally time-based rather than activity based. Our participants recommend for more intelligent progress tracking so that personalized reminders which based on individuals’ completion of task could be delivered by the system. While we foresee this leading to greater administrative challenges for MOOC instructors, we also see the potential for the same set of triggers to be used for adaptive delivery or adaptive assessment.

5.8 Leverage on M-Learning

Finally, our research participants recommend that MOOC providers leverage on the potential of mobile learning and the pervasiveness of mobile devices among students. This could be the provision of mobile
application linking learners to the MOOC platform and courses, syncing course dates to the calendar in learner’s mobile devices; or taking advantage of existing collaboration tools such as group messages.

6. Limitations of Research

A limitation of this study is its possible lack of generalizability. Though the sample size is large enough for such qualitative study, the findings are typically relevant to the specific group of learners under investigation, with its own characteristics. While this study is academic staff-specific, our goal is to share recommendations on the development of MOOCs for those who are interested to offer courses via this mode. We recommend for further studies to be done on more samples of MOOC learners to gain more objective inputs.

7. Conclusion

There are many factors that motivated learners to sign up for learning via MOOCs and successful completion for their choice of MOOC subjects. Instructors could promote individual interest by (1) providing learners with opportunities to have control over their learning (2) relating the usefulness of content to achieve their goals (3) creating a warm and personalized presence to help learners feel connected and engaged (Osborne et al., 2007).

The growth of online learning options continues to increase and will have an impact on the shape of higher education. Learners in the 21st century want learning experiences that support participation and social interaction that will connect them with their peers.

In this study, strategies on how MOOCs could be made more engaging and adaptive to learners with different learning needs and styles were made. We hope that these strategies and methods could help instructors design an online learning environment that meets the needs and learning behavior of students in the 21st century. By integrating support, instructors could help learners to have greater insights about effective time management, prioritizing and stay motivated throughout the course. In addition, they could help learners with different learning needs to identify a pathway for successful online learning.

References


Effects of Gender Differences and Learning Performance within Residence Energy Saving Game-based Inquiry Playing

Ugyen DORJI*, Patcharin PANJABUREEb & Niwat SRISAWASDIc

aGelephu Higher Secondary School, Bhutan
bInstitute for Innovative Learning, Mahidol University, Thailand
cFaculty of Education, Khon Kaen University, Thailand
*panjaburee_p@hotmail.com

Abstract: Energy Saving in Residence is one of the important topics in Energy Education, whereby students are required to understand the factors of energy consumption and conservation. Several researchers have studied on how to use digital games to improve understanding the topic. However, not every student understands is improved by using these kinds of learning assistance. Therefore, a study on effect of gender differences and learning performance within energy saving games have become a need to address the issue. This paper has attempted to develop a digital game based on inquiry-based learning called Residence Energy Saving Battle (RES-battle). Further, the paper examines the effect of gender difference in students’ learning performance and the attitude towards the RES-battle. The result shows that students’ learning performances on energy consumption and conservation that significantly improved after they participated in the RES-battle. The result also revealed that the RES-battle can decrease the difference between female and male learning performances reasonably and attitude toward the RES-battle. From the findings, this paper has implication for the development of students’ residence energy consumption learning within the digital game.

Keywords: Game-based learning, pedagogy, gender study, energy education

1. Introduction

Electricity, a one of longest-used demand, is widely used for lighting or heating purposes and to make electrical equipment works. Due to widespread use, most of consumers are careless in its management. The amount of electrical energy usage occurs not only from residence budget and size but also from residents’ behavior in using electrical equipment (Moll, Noorman, Kok, Engström, Throne-Holst, and Clark, 2005). In the past decade, several researchers attempted to provide the way to decrease the electricity usage in each residence. For example, Abrahamse, Steg, Vlek, and Rothengatter (2005) focused on altering the residents’ behavior in the use of electrical equipment in appropriate way. Maharaj-Sharma (2012) suggested that student energy learning is need to apply and link to outside school and daily life. Consequently, learning about energy consumption and conservation in school science has become important in current research.

Although energy consumption and conservation is one of important physics concept in school science. Generally, this topic is a part of physics teaching and learning about electricity. In the conventional physics class, student learned electricity as attribute of electronic and consumption and conservation of energy in a form of employed energy calculation, but not as procedures as it is recognized in contemporary physics. As such, it is an instructional challenge to motivate student learning the concept meaningfully. Several methods and strategies have been used to make the concept of energy consumption and conservation more meaningful for understanding to the students and for encouraging them linking the concept to daily life. For example, Chiu, Chou, and Liu (2002) suggested that although using analogies or metacognition might help the learning process more permanent, but there is no significant conceptual development by these approaches. Slotta and Chi (2006) recommended that students can insight what concepts from their own understanding. Thus, teaching
students to use electric energy efficiently in their daily life to encourage them to construct conceptual understanding of energy consumption and conservation by themselves might be useful way for saving energy in residence. In other words, digital interactive learning could be a useful way for conceptual learning of energy consumption and conservation and for simulating practices of saving energy in residence.

On the other hand, digital interactive learning is generally developed as multimedia learning units with simulations and games in many disciplines such as in mathematics, science, engineering, humanities, and social sciences (Cai, Lu, Zheng, and Li, 2006; Eck, 2006). Recent research attempted to purpose educational computer game which has advantage in students’ habits and interests (Gee, 2006; Prensky, 2007) and indicated that the use of digital game is another useful method to improve energy consumption and conservation learning (Yang, Chien, and Liu, 2012). Playing educational computer game requires prior knowledge or pre-existing learning experience that helps students to apply the knowledge to make decision related to realistic situations (Papastergiou, 2009) and can be considered as a learning tool for teaching the factual information as well as worksheet activities (Spraggins and Rowsey, 1986). Therefore, developing a digital learning of energy consumption and conservation that integrates game-based learning technologies and pedagogy into learning of energy consumption and conservation may be a useful way to improve students’ learning performance on the topic. However, the successful usage of pedagogy-driven digital game depends on the digital game, the learning strategies, and human factors. Among various human factors, gender difference play an important role when playing digital game affecting on learning performance (Paraskeva, Mysirlaki, and Papagianni, 2010).

In summary, pedagogy-driven digital game, gender differences, and conceptual learning performance are critical to learning saving energy in residence in which students are required to understand the energy consumption and conservation meaningfully. In this vein, the aim of this study is to examine the effects of gender differences on students’ conceptual learning performance when playing the Residence Energy Saving Battle (RES-battle) digital game and attitudes toward the RES-battle.

2. Background and Motivation

In Bhutanese education system, science subject is introduced from primary level onwards and when it reaches in higher secondary level it is segregated into three disciplines; namely Physics, Chemistry and Biology. The aim of having this is to make the curriculum more relevant to the need of the society or localization and to bring a shift in the teaching style from teacher centered to learner-centered. However, in many schools in Bhutan teacher-centered instruction is a common practice. Most students in Bhutan lack understanding of concepts and they learn by memorization. Many science teachers feel and our personal experiences from the past has found that the answers that students write in the exams are just regurgitation of what they have learnt by memorization. Any twist or rephrasing of the questions made the students difficult or unable to answer the questions.

In recent years, the education system supplied hundreds of computers all over the nation under the project “Chiphen Rigphel Project” (Information Communication Technology-ICT Project) with grant from government of India. The objective of the project is to equip every student with IT literacy and improve science education. Now the system is stressing more on the practical oriented learning that enable students to acquire the skills that can be applied in the practical context. So, the system has observed that digital learning as one of the ways to achieve that goal. Therefore, there is strong need for teachers to change the instructional practices from teacher centeredness to child-centeredness by practicing effective teaching strategies with integration of digital learning like computer games, animations, simulations, and more, particularly in learning abstract concepts in science curriculum, especially on topic energy consumption and conservation.

Energy education has become very important element to educate students on the basic energy concept on daily electrical energy consumption and conservation. It help students to identify basic factors on which energy consumption depend and provide useful ways for reducing daily energy consumption both in and outside of their home and school. Integrating appropriate learning approaches for teaching the concept is very important for any effective learning process. Thus, introducing the concept of energy education into the ways the students understand is crucial for any educator (Gustafson and Branch, 2002). Moreover, one of effective learning strategies for making students as
active learner is inquiry-based learning approach (Kubicek, 2005). It requires teachers to provide the opportunity for students to observe, gather, analyze, and interpret data which students learn from their findings, explanation, predictions, and communications with peers, which led students to construct their own conceptual knowledge (Krajcik and Blumenfeld, 2006; Kuhn, Black, Keselman, and Kaplan, 2000).

Therefore, in this study, the inquiry-based learning approach was chosen to drive a digital game-based learning to assist students exploring factors related energy consumption, by simulating using electrical appliances for certain durations led to saving energy in residences. These features of the game might help students to improve conceptual learning performance on the topic of energy consumption and conservation. Moreover, when playing digital game, students’ gender difference factors play might effect on their learning performance. Consequently, there is a need to examine how the RES-battle affect students’ learning performance, and how gender differences affect students’ learning performance of conceptual knowledge on energy consumption and conservation. To this end, this study emphasizes on whether gender differences affect students’ learning performance by taking the RES-battle.

3. The Digital Game based on Inquiry-based Learning Approach: Residence Energy Saving Battle (RES-battle)

With the benefit of game-based learning in motivating students learning and inquiry-based learning approach in encouraging students explore and construct conceptual knowledge on energy consumption and conservation, this study developed Residence Energy Saving Battle (RES-battle) in which the inquiry-based learning approach drives activities in computer-based game playing. The RES-battle is designed in line with practical situations of energy consumption in our daily life. To save money and minimize energy consumption, these help the students in developing sense of awareness on energy conservation. The procedure of RES-battle is following steps:

- **Step 1:** The objectives of the RES-battle are introduced to the students, such as the students should be able to know how energy consumption is calculated in term of money and also how can they apply those factors in monitoring power bill by playing the RES-battle.

- **Step 2:** The rules and basic functions of the RES-battle are demonstrated to the students, such as the RES-battle is designed with scenarios of home comprising electrical appliances that are commonly operated. It is divided into two different-difficulty levels in which the students can go to the second level by passing the first level. Each level has 30 playing-time second. To get into another level, the students should have enough income/budget earned from current stage. In doing, the students should hunt/collect hidden coins which depend upon wattage of those appliances in that room. The students have to take cautious decisions on selecting/clicking on appliances to get those coins because every time they collect coins, it activates the appliance which consumes energy, and they have to pay at the end.

![Figure 1. Illustrate playing the Residence Energy Saving Battle (RES-battle) in the first level](image-url)
• **Step 3:** While playing the RES-battle, whatever coins are accumulated and how much energy has consumed are automatically calculated and displayed in graphical format at the side of the game screen as shown in Figure 1. These help students to apply their theoretical knowledge, skills and strategies to make decisions in their assigned role which promotes the understanding of concept. The wattage of each appliance will be shown randomly at certain interval. The value of hidden coins in each appliance remains constant whether it is chosen at the beginning or at last. In this step, the students are encouraged to inquire factors of energy consumption which are wattage of the appliance and duration of usage.

![Image of Energy Consumed Graph](image)

The graph shows the rate of energy consumption of the selected appliances. The result shows how much you save money.

**Figure 2. Illustrate results of playing the Residence Energy Saving Battle (RES-battle) in the first level**

• **Step 4:** At the end of 30 second, the RES-battle automatically is stopped and displays “Time Up”, after that simulated graph for energy consumption will be displayed for those appliances that the students operated during the game as shown in Figure 2. Based on that, cost for energy consumed is calculated and then saving of the students is shown \((\text{your saving} = \text{income} - \text{cost of energy})\). If the saving is positive, the player can continue the game in the second level with more difficulty as shown in Figure 3, otherwise they cannot go further or can play again as shown in Figure 2.

4. **Research Methodology**

To investigate the effectiveness of the RES-battle, we investigate whether the RES-battle improve students’ learning performance on energy consumption and conservation; result on this will led us to examine whether gender differences affect students’ learning performance by taking the RES-battle and whether gender differences affect students’ attitudes toward the RES-battle.

This study was designed by one group pre-posttest design. A total of 68 tenth graders of secondary school students in eastern Bhutan were recruited in this study. They participated in the RES-battle individually, and discussed the situations in the RES-battle with their peer to construct their own knowledge of factors of energy consumption and conservation. Before playing in the RES-battle, the students were asked to take pre-conceptual test of the topic. After finishing learning activities on the RES-battle, they were asked to take post-conceptual test. Moreover, they are asked to respond the attitude questionnaire to clarify the degree of their attitude toward the RES-battle.

To examine whether the RES-battle improve students’ learning performance on energy consumption and conservation, pre- and post-conceptual tests were used as the research tools. The validity of the tests were determined by three experienced teachers teaching same subject. Each test contained 20 multiple-choice items, and one point was scored for each correct answer; therefore, the total score of the each test was 20.
To investigate students’ attitudes toward the RES-battle, the attitude question adopted from Subba (2011) was used in this study. It consisted of 18 items, which were categorized into three categories: interest, participation, and satisfaction. This questionnaire was measuring using a 5-points Likert scale in which for “Interest” and “Participation” category, levels 5 represents for “Strongly agree”, 4 for “Agree”, 3 for “Neutral”, 2 for “Disagree” and 1 for “Strongly disagree”; whereas in “Satisfaction” category, 5 represents for “Extremely Satisfaction”, 4 for “High Satisfaction”, 3 for “Moderate Satisfaction”, 2 for “Low Satisfaction” and 1 for “Least Satisfaction”.

Figure 3. Illustrate success of the first level and move to the second level with more difficulty

5. Results

5.1 Students’ conceptual learning performance

5.1.1 Overall conceptual learning performance

Table 1 shows the results of the conceptual test, in term of Mean and Standard Deviation (SD), between pre- and post-test. The result shows that there was a significant difference between the pre- and post-test on students’ conceptual learning of energy consumption and conservation ($t = 13.887, p = .000$). It indicates that the students performed better in the post-test than in the pre-test; implying that students’ conceptual learning were significantly improved after participating in the RES-battle. As shown in Table 1, it seems obviously that the SD in the post-test was little increased from the SD in the pre-test. This result indicates that there is different amount of conceptual knowledge gain after
taking the RES-battle; some student has high gain conceptual knowledge and some student has low
gain conceptual knowledge. Thus, another data analysis was conducted to clarify the conceptual
knowledge gain in the next section.

Table 1: Conceptual learning performance between pre- and post-test

<table>
<thead>
<tr>
<th>Test</th>
<th>N</th>
<th>Mean</th>
<th>SD</th>
<th>t</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre-test</td>
<td>68</td>
<td>7.56</td>
<td>2.27</td>
<td>13.887**</td>
</tr>
<tr>
<td>Post-test</td>
<td>68</td>
<td>13.35</td>
<td>3.03</td>
<td></td>
</tr>
</tbody>
</table>

**p < 0.01

5.1.2 Conceptual learning progression

To clarify the conceptual knowledge gain of the students after taking the RES-battle, this study
employed normalized gain of Hake (1998) which defined the \(<g>\) as “High gain, \(<g> \geq 0.7\”,
“Medium gain, 0.7 > \(<g> \leq 0.3\” , and “Low gain, \(<g> < 0.3\” . From sixty-eight students’ conceptual
pre- and conceptual post-test scores, the results show that there are 13, 36, and 19 students for high,
medium, and low gains respectively. Thus, this results indicates that most of students gains conceptual
knowledge of energy consumption and conservation moderately after taking the RES-battle. For
overall result, as shown in Table 2, the conceptual score of pre- and post-test, the \(<g>\) is 0.47
indicating that the students have conceptual learning progression of their learning by gaining better
conceptual knowledge after playing the RES-battle.

Table 2: Conceptual learning progression by the average normalized gain \(<g>\)

<table>
<thead>
<tr>
<th>Conceptual test (Total score = 20)</th>
<th>N</th>
<th>Mean</th>
<th>S.D.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre-test</td>
<td>68</td>
<td>7.56</td>
<td>2.27</td>
</tr>
<tr>
<td>Post-test</td>
<td>68</td>
<td>13.35</td>
<td>3.03</td>
</tr>
<tr>
<td>(&lt;g&gt;)</td>
<td></td>
<td>0.47</td>
<td></td>
</tr>
</tbody>
</table>

5.2 Students’ attitudes toward the Residence Energy Saving Battle (RES-battle)

Table 3 shows the descriptive statistics of the students’ attitude toward the RES-battle. Overall
analysis from the questionnaires reveals that the students have rated “High Agree” for the RES-battle
for being able to develop interest in learning the concept of energy consumption and conservation in
residence. Moreover, the students respond that they were “Highest Agree” towards the RES-battle in
which using the inquiry-based learning drives learning activities on the topic. Because the RES-battle
encouraged them to explore and construct conceptual knowledge. In addition, the students rated that
they have “High Satisfaction” in learning concept of energy consumption and conservation through
the RES-battle.

Table 3: Students’ attitude degree after taking the RES-battle

<table>
<thead>
<tr>
<th>Attitude Aspect (N=65)</th>
<th>Mean</th>
<th>SD</th>
<th>Interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Interest</strong></td>
<td>1. I enjoy learning very much with the RES-battle.</td>
<td>4.33</td>
<td>0.63</td>
</tr>
<tr>
<td></td>
<td>2. I become more curious and observant in the class when the lesson is integrated with the RES-battle.</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Participation</strong></td>
<td>3. I enjoy participating in class activities when the lessons are taught using the RES-battle.</td>
<td>4.45</td>
<td>0.57</td>
</tr>
<tr>
<td></td>
<td>4. The RES-battle make me more attentive in the class.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
5. Integration of the RES-battle in the lesson promotes better interaction amongst friends and teachers.

6. It is easier for me to understand the content with the RES-battle.

7. I get learning satisfaction when I learn the lesson with the RES-battle.

8. The RES-battle in the lesson helps me to develop confidence in learning electrical energy calculation.

9. I found the RES-battle useful in visualizing the concepts.

10. The RES-battle in learning helps me to think and analyze the real things in world.

11. It helps me to develop the relevance between the course and real world situations.

12. The RES-battle allows me to develop skills needed in the real world.

13. I like the way the teacher uses the RES-battle to teach energy consumption by various household appliance lesson.

14. The use of the RES-battle in the lesson helps me to build confidence in understanding the concept of energy consumption clearly.

15. I like the RES-battle in learning electrical energy consumption and conservation sessions because it enables me to learn faster.

16. I gain confidence when I learn the lesson using the RES-battle.

17. I like electrical energy lessons with the RES-battle because the lessons are interesting, informative and help to visualize the abstract concepts of energy better.

18. I like the RES-battle integration in all the subjects to help enhance our critical thinking.

5.3 Gender Differences

5.3.1 Gender differences in overall conceptual learning performance

Table 4 shows the results of students’ conceptual learning performance, in term of mean and Standard Deviation (SD), between females and males for conceptual learning performance. The results of the pre-test and post-test show that there were no significant difference between females and males for their conceptual learning performance. In other words, conceptual learning performance between females and males had no difference in the post-test. It indicates that females and males improved at a faster rate and were not statistically different in the end of the RES-battle.

Table 4: Conceptual learning performance between females and males in the pre- and post-test

<table>
<thead>
<tr>
<th>Test</th>
<th>Gender</th>
<th>N</th>
<th>Mean</th>
<th>SD</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre-test</td>
<td>Female</td>
<td>38</td>
<td>7.32</td>
<td>2.28</td>
<td>.994</td>
<td>.162</td>
</tr>
<tr>
<td>Pre-test</td>
<td>Male</td>
<td>30</td>
<td>7.87</td>
<td>2.25</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Post-test</td>
<td>Female</td>
<td>38</td>
<td>12.95</td>
<td>2.94</td>
<td>1.249</td>
<td>.108</td>
</tr>
<tr>
<td>Post-test</td>
<td>Male</td>
<td>30</td>
<td>13.87</td>
<td>3.10</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
5.3.2 *Gender differences in different sizes of conceptual learning gain*

To examine the gender difference for the high, medium, and low gain size of conceptual knowledge, analyses were undertaken between boys and girls in the pre- and post-test for the three groups in different sizes of conceptual learning gain. Table 5 shows the results of these analyses. The results show that there were no significant differences between females and males in the pre- and post-test of the three gain size groups. These findings reveal that females and males improved but were not statistically different in the end of the RES-battle.

Table 5: Performance of the high, medium, low gain size between females and males in the pre- and post-test

<table>
<thead>
<tr>
<th>Size</th>
<th>Test</th>
<th>Gender</th>
<th>N</th>
<th>Mean</th>
<th>SD</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>High</td>
<td>Pre-test</td>
<td>Female</td>
<td>7</td>
<td>5.57</td>
<td>2.99</td>
<td>1.543</td>
<td>.075</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Male</td>
<td>6</td>
<td>8.17</td>
<td>3.06</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Post-test</td>
<td>Female</td>
<td>7</td>
<td>17.00</td>
<td>.816</td>
<td>1.721</td>
<td>.056</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Male</td>
<td>6</td>
<td>17.67</td>
<td>.516</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Medium</td>
<td>Pre-test</td>
<td>Female</td>
<td>18</td>
<td>7.78</td>
<td>2.102</td>
<td>.156</td>
<td>.438</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Male</td>
<td>18</td>
<td>7.67</td>
<td>2.169</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Post-test</td>
<td>Female</td>
<td>18</td>
<td>13.61</td>
<td>1.787</td>
<td>.494</td>
<td>.312</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Male</td>
<td>18</td>
<td>13.94</td>
<td>2.235</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low</td>
<td>Pre-test</td>
<td>Female</td>
<td>13</td>
<td>7.62</td>
<td>1.758</td>
<td>.616</td>
<td>.573</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Male</td>
<td>6</td>
<td>8.17</td>
<td>1.941</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Post-test</td>
<td>Female</td>
<td>13</td>
<td>9.85</td>
<td>1.068</td>
<td>.023</td>
<td>.491</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Male</td>
<td>6</td>
<td>9.83</td>
<td>1.329</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

5.3.3 *Gender differences in attitude toward the Residence Energy Saving Battle (RES-battle)*

Table 6 shows the results of students’ attitude toward the RES-battle, in terms of mean and Standard Deviation (SD), between females and males. The results of the three attitude aspects show that there were no significant differences between females and males for their attitude. In other words, attitude toward the RES-battle between females and males were not different. It indicates that females and males satisfied equally in the end of the RES-battle.

Table 6: Attitudes between females and males after taking the RES-battle

<table>
<thead>
<tr>
<th>Attitude Aspect</th>
<th>Gender</th>
<th>N</th>
<th>Mean</th>
<th>SD</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interest</td>
<td>Female</td>
<td>37</td>
<td>4.38</td>
<td>.57</td>
<td>.655</td>
<td>.256</td>
</tr>
<tr>
<td></td>
<td>Male</td>
<td>29</td>
<td>4.27</td>
<td>.70</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Participation</td>
<td>Female</td>
<td>37</td>
<td>4.50</td>
<td>.55</td>
<td>.748</td>
<td>.223</td>
</tr>
<tr>
<td></td>
<td>Male</td>
<td>29</td>
<td>4.39</td>
<td>.61</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Satisfaction</td>
<td>Female</td>
<td>37</td>
<td>4.33</td>
<td>.40</td>
<td>1.566</td>
<td>.061</td>
</tr>
<tr>
<td></td>
<td>Male</td>
<td>29</td>
<td>4.50</td>
<td>.43</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

6. Conclusion and Implications

This study examined the effectiveness of a digital game in which the inquiry-based learning approach drives activities within the game. While playing the game, students were encouraged to explore elements in the game, discuss what they found with their peer, and construct their own knowledge. Although Gee (2007) and Unlusoy, de Haan, Leseman, and van Kruistum (2010) revealed that males interests in digital games is more than females do. But the findings from this study highlighted that there were no difference in the conceptual tests and attitude towards the developed game between females and males. Moreover, when the participants were categorized based on their conceptual knowledge gain size into high, medium, and low gain size, there were also no difference in the
conceptual tests, and attitude toward the developed game between females and males. These findings could imply that the difference between females and males was reduced after taking the developed game. In other words, the developed game enabled the students who gained conceptual knowledge in different size to make significant conceptual learning improvement.

Therefore, to decrease the gap between females and males on conceptual learning performance when using the digital game-based learning, there is a need to develop the game that can provide the opportunities for interaction on game screen and also with other students. The road of game playing needs to accommodate the encoding and decoding of graphics through the game and the elements that promote students inquired evidence for constructing the knowledge.

Acknowledgement

This research project is supported by Thailand International Cooperation and Development Agency (TICA) scholarship.

References


A Blended Learning Environment in Chemistry for Promoting Conceptual Comprehension: A Journey to Target Students' Misconceptions

Sumarin NIROJ* & Niwat SRISAWASDIb*

aScience and Technology Education Program, Faculty of Education, Khon Kaen University, Thailand
bDivision of Science, Mathematics, and Technology Education, Faculty of Education, Khon Kaen University, Thailand

*niwsri@kku.ac.th

Abstract: Blended learning has become a new form of learning and teaching in science education. Researchers have indicated that the blended learning could create more meaningful in learning, support deep-level understanding, and contribute to better learning achievement in the exam. The present article presents a blended learning environment, combination of computer-based interactive lecture demonstration (CBILD) and web-based inquiry science environment (WISE), for chemistry learning of state of matter and phase change. To create the blended environment, 50 twelve-graders were recruited in an investigation of prior knowledge by challenging their epistemological and ontological belief of the chemistry concepts. As such, a series of two-tier conceptual items were administered and then the students' unscientific conceptions about state of matter and phase change were extracted as an account. The finding showed that they hold various patterns of misconceptions covering the effects of pressure and temperature on arrangement of particle and phase change, and shape of molecule during phase change. Moreover, all of them have no scientific conception that plasma is a state of matter. Many misconceptions seem to due to the fact that they could not realize interaction between observable and unobservable level of chemical phenomena because of its complexity and abstraction. To make chemistry more accessible and meaningful for student learning, CBILD learning process was design in emphasizing macro-level representation and WISE was design to support sub-micro-level and symbolic representations in chemistry as face-to-face and on-line learning experience respectively. The learning process of blended learning environment was illustrated and described in a sequence for implementing in chemistry classroom. This could be an implication for researchers and teachers how to create blended learning environment which could improve teaching and learning strategy into a new form for science classroom, and it might enhance the change of student's misconceptions and their mental model development in science.

Keywords: Blended learning, interactive lecture demonstration, web-based learning, microcomputer-based laboratory, chemistry education

1. Introduction

Chemistry is a fundamental science which is abstract and complex by its nature. Due with its nature, students lack of transfer what they learned, e.g. concepts, to real-world problems and everyday life, and they give no meaning to what they have learned (Gilbert et al., 2002; Gilbert, 2006). Although, the chemistry learning activities attempted to link the subject matter with how the world works, the students still have numerous learning difficulties and misconceptions on the subject. Moreover, they merely link their own existing ideas to the new concepts leading to fragmented and fractured understanding (Gilbert and Boulter, 2000). Such misconceptions, especially on the topic state of matter and phase change, are due to the fact that the students could not distinguish between macroscopic and sub-microscopic explanations. The students have also difficulty linking observable phenomena to molecular level interaction (Chang and Linn, 2013). As such, many studies suggested that the molecular visualizations
could help students integrate observable, molecular, and symbolic aspects of chemical change (Russell and Kozma, 2005).

With the benefit of computers and technology in science education community, the contemporary technology-based approach has been used to enhance students’ conceptual understanding (Vreman-de Olde et al., 2013; Srisawasdi and Kroothkeaw, 2014). Moreover, the inquiry-based learning strategy supported by Information and Communication Technology (ICT) has known as an effective pedagogical approach in science learning. ICT can also enable new ways of education to deliver knowledge directly from teachers to students and the learning can take place anywhere and anytime. Recently, there are many computer-supported inquiry-based science learning environments such as WISE, Co-Lab, Inquiry Island, and nQuire (Sun et al., 2013). The instructions that teachers implement have been changed by conducting the new learning paradigm and innovative educational tools. Active learning and knowledge sharing have been replacing traditional teacher-centered lectures, due to their ability to integrate ICT into the curriculum and their emphasis on meaningful learning (Yen and Lee, 2011).

Blended (hybrid, mixed-mode) learning is referred as learning combination of the classical teaching in classroom (face-to-face learning) and teaching assisted by contemporary technologies. Thorn (2003) described blended learning as a way of meeting the challenges of tailoring learning and development to the needs of individuals by integrating the innovative and technological advances in online learning environment. Moreover, Yapici and Akbayin (2012) revealed that the students who participated in blended learning model could achieve learning outcomes in biology course than those who participated in the traditional teaching method significantly. In additions, they also have positive attitude toward science if the learning environment had incorporated by internet access (Yapici and Akbayin, 2012). As such, the blended learning could be used to overcome the difficulties of practical science.

With the importance of enhancing students’ conceptual understanding in science concepts and considering benefits of blended learning environment as aforementioned, the researchers aim to develop an effective blended learning environment for chemistry learning of state of matter and phase change regarding students' specific misconceptions. As such, investigation of the misconceptions before designing of science learning environment could be a strategic way for research and development in science and technology education.

2. Blended Learning Environment

Lecture demonstration method is usually an integral part of a university education. This method helps students to visualize the material through students’ understanding of the subject. Zimrot and Ashkenazi (2007) indicated that interactive lecture demonstrations (ILD) could be used to make more accessible of scientific conceptual understanding because of potential source of anomalous data that can trigger cognitive process of conceptual change. Moreover, ILDs are fun to do and provide concrete examples of abstract concepts. Nevertheless, using only the traditional approach to education, where the transfer of knowledge is achieved mostly by lecturing, has a number of shortcomings, because the students are not motivated enough to acquire knowledge actively (Hoic-Bozic, Mornar and Boticki, 2009; Burnham, 2001). To cope with this issue, the computer-based interactive lecture demonstration (CBILD) could be a novel of instruction to encouraging students learning and is one component of blended learning environment in this study.

Moreover, to enhance knowledge integration, the use of computer visualizations in inquiry activities can enhance students’ conceptual understanding of molecular level interactions. web-based inquiry science environment (WISE), which is one of web-based learning environment and is one component of blended learning environment in this study, was developed based on the knowledge integration perspective that help students develop a more cohesive, coherent, and thoughtful account of scientific phenomena (Linn et al, 2003). WISE is a powerful, research-based online platform for designing, developing, and implementing science inquiry activities. Unique features and benefits of WISE consists library of free, assessments aligned with instruction, interactive visualizations and simulations, embedded prompts for reflection and collaboration, instructional support for diverse
learners, teacher feedback and guidance tools, powerful authoring and customization tools library of free, and classroom-tested projects.

Therefore, in this study, the blended learning environment refers to a combination of computer-based interactive lecture demonstration (CBILD) and web-based inquiry science environment (WISE). In this blended learning environment, the learning sequence to facilitate the construction of conceptual understanding and induce the change of misconceptions on the topic state of matter and phase change has been designed and presented in this paper.

3. The Exploration of Students’ Misconceptions

In this study, the researchers conducted an exploration to identify students’ common misconceptions on chemistry topic of state of matter and phase change. The findings of the exploration provided us as a basis in order to design and create the blended learning environment by combining CBILD and WISE as a novel learning experience for chemistry learning.

3.1 Participants

To explore misconceptions on the topic, fifty twelve-grade students, age ranging from 17 to 18 years in a local public school at the Northeastern region of Thailand were recruited in to the study. They already completed a regular chemistry class and they were taught about the state of matter and phase change before participated in this study.

3.2 Research Instrument and Data Analysis

As aforementioned, 20-item two-tier conceptual test targeting chemistry concepts of state of matter and phase change was used to investigate students’ common misconceptions. The conceptual test was adapted from a published research instrument, Particulate Nature of Matter (ParNoMa) (Bridle and Yezierski, 2011). Some of the two-tier conceptual items were constructed by the researchers, and the questions were developed specifically for the concept of phase change of gas to plasma concepts. This topic was chosen after an extensive literature review, which reported that there was no study about student’s conceptual understanding in plasma state of matter before. The two-tier conceptual item contains a first tier of multiple choices associated with the main question and a second tier of open-ended reasoning. The two-tier conceptual test was reviewed by three experts for identifying construct and communication validity. The respondents were asked to complete the test within 60 minutes. In this study, the researchers have analyzed and interpreted the respondents’ answers into four categories of conception: scientific conception (SC), incomplete conception (IC), misconception (MC), and no conception (NC). The misconceptions identified by this study were used in the development of blended learning environment for chemistry learning of state of matter and phase change.

3.3 Results of Students’ Misconceptions on State of Matter and Phase change

The two main topics in this study were state of matter and phase change. These topics consist of many fundamental concepts in chemistry. The concepts were: (i) effect of pressure on the arrangement of particle; (ii) particle movement; and (iii) physical properties of matter, i.e. solid, liquid and gas, and plasma, a new-defined state of matter. After the survey, the researchers found that students hold various patterns of misconceptions in the first topics. In this paper, the researchers reported only students’ misconceptions on the effect of pressure on arrangement of particle that most students illustrated especially with many misconceptions. In the second topic, most of students’ misconceptions were in place with chemistry phenomena of bubble formed during boiling, phase change of gas to plasma, shape of molecule and the phase change, and the relationship between temperature and phase change. Table 1 displays percentage of students’ misconceptions on state of matter and phase change.
Table 1: Students’ misconceptions about state of matter and phase change.

<table>
<thead>
<tr>
<th>Topic</th>
<th>Concept</th>
<th>Specific types of misconception (MC)</th>
<th>Frequency (%)</th>
</tr>
</thead>
</table>
| State of matter                | Effect of pressure on particle arrangement | When pressure of carbon dioxide decreased:  
  - No effect on space and size of its molecules  
  - Decreasing of molecule attachment and getting smaller in size  
  - More diffusion of substances  
  - Increasing of space among molecules and getting larger in size  
  - Increasing of numbers of molecule  
  - Decreasing of space among molecules but no effect on size of molecule | 12 8 2 20 12 24 |
|                               |                                | **Total**                                                                                          | 78            |
| Phase Change                   | Bubbles formed during boiling  | The bubbles were produced because:  
  - Making change by heat  
  The bubbles come from:  
  - Air  
  - Decomposing of water into hydrogen atoms and oxygen atoms  
  - Liquid pressure | 20 8 22 2 |
|                               |                                | **Total**                                                                                          | 52            |
| Phase change of gas to plasma  |                                | When gas reach to high temperature around 10,000 Celsius:  
  - Decomposing into hydrogen molecules and oxygen molecules  
  - Decomposing into \( H^+ \) and \( OH^- \)  
  - Decomposing to the end of nothing  
  - Nothing happened | 36 14 14 14 |
|                               |                                | **Total**                                                                                          | 78            |
| Shape of molecule and phase change |                                | The condensation causes:  
  - Larger size of water molecule  
  - Larger size of water molecule because of its combination  
  The freezing causes:  
  - Larger size of water molecule because of its combination | 13 12 17 |
|                               |                                | **Total**                                                                                          | 46            |
| The relationship between temperature and phase change | When ice was heated:  
  - Changing to be water and stream state respectively | **Total**                                                                                          | 66            |

4. Discussion

This pilot study reports students’ misconceptions on state of matter and phase change and the design of blended learning environment combining CBILD with WISE based on the findings of students’ misconceptions. The result showed some patterns of students’ misconceptions about state of matter and phase change. This could be discussed that students hold with various misconceptions in the same
content, it shows difference of understanding although they were provided in the same learning context in classroom. Many misconceptions are due to the fact that students do not distinguish between macroscopic and submicroscopic explanations and students have also difficulty linking observable phenomena to molecular level interaction (Chang and Linn, 2013). Almost all students could not explain in term of molecular level, some can answer correctly in multiple choice part of two-tier test but cannot give a reason of their answer or give it by not corresponding. Most chemistry courses require improving students’ understanding by adding scientific concepts; however, students often add concepts without integrating them with their other related concepts, which leads to fragmented and fractured understanding (Gilbert and Boulter, 2000). Hence, the researchers can conclude that knowledge integration affects conceptual understanding.

To enhance scientific conceptual understanding the instructor needs to provide instruction for supporting the knowledge integration. The blended combination of CBILD and WISE was designed to address students’ misconceptions by supporting knowledge integration and conceptual understanding to make science accessible.

5. The Design of Blended Learning Environment to Address Students’ Misconceptions about State of Matter and Phase Change

The blended learning was considered learning environment to address students’ misconceptions about state of matter and phase change. Face-to face learning takes place in the conventional science classroom using the computer-based demonstration provided by teacher in front of classroom via projector. The computer-based demonstration may be science experiment or video presentation that dependent on the content and accessible learning materials. For example, the experiment of plasma cannot be demonstrated or performed its experiment in classroom because it is very harmful and instrument also rare in laboratory. During learning period, students took one hour for CBILD in conventional classroom and then two hours for WISE in computer room. Teacher provided student CBILD strategy to make science accessible by empirical evident that is observable phenomena and lead them to distinguish macroscopic explanation. Student were further conducted to investigate unobservable phenomena in the molecular level by using molecular visualization supporting sub-microscopic explanation related to macroscopic and symbolic explanation. When they gain knowledge by related concepts, they will can link three explanations together and result in knowledge integration that becomes conceptual understanding, as shown in Figure 1.

![Figure 1. Features of blended learning environment to facilitate students’ conceptual understanding in chemistry.](image-url)
5.1 An Example of CBILD on the Relationship between Temperature and Phase Change

In this part, the researchers would like to present emerging the new technology into science classroom to make science to support students’ conceptual understanding and meaningful learning about scientific content from empirical evident. The Vernier Software and Technology is conducted to demonstration of experiment, us sensors such as temperature probe to detect temperature during phase change process of water, and also use a gas pressure sensor to measure vapor pressure of liquid for study the effects of temperature and type of liquid on vapor pressure of liquid. These sensors can connect with computer supported by LabQuest Mini and Logger Pro software to acquire real time collection and analysis data, and directly display the result by giving a real-time graph. The researchers use interactive lecture demonstration (ILD) strategy to provide instruction. They help students to visualize the material that is being discussed in class and are frequently described by students as an important component of their understanding of the subject. In this method, the students are asked to predict the outcome of a demonstration and write down their prediction and explanation, and therefore commit to an explicit model. Peer discussion follows, with the students discussing their predictions in small groups—again, they have to address their existing models explicitly (see Table 2).

Table 2: An example of learning process in CBILD classroom.

<table>
<thead>
<tr>
<th>Components</th>
<th>Description of learning process</th>
<th>Examples of learning activity</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Confront</td>
<td>Teacher engages student by thinking and giving examples about phase change phenomena in daily life and then provides questions about relationship between temperature and phase change.</td>
<td><img src="image1" alt="Image" /></td>
</tr>
<tr>
<td>2. Predict</td>
<td>Teacher conducts students to predict how does temperature change if ice is heated and become water and finally to vapor, respectively, and then water become ice after cooling.</td>
<td><img src="image2" alt="Image" /></td>
</tr>
<tr>
<td>3. Experience</td>
<td>Teacher demonstrates change of temperature during phase change by using the Vernier Software and Technology to real time data collection and analysis.</td>
<td><img src="image3" alt="Image" /></td>
</tr>
<tr>
<td>4. Reflect</td>
<td>Teacher allows students to compare the results of experiment and their previous prediction and engages them to discuss on the confront phenomena Teacher induces student into formative assessment question or statement related the phase change phenomena.</td>
<td><img src="image4" alt="Image" /></td>
</tr>
</tbody>
</table>
5. Transfer

- Teacher encourages student to explain another physical phenomenon based on the phase change concept.

5.2 An Example of WISE on State of Matter

Students will use the visualization, resources from The Concord Consortium, PhET, Molecular Workbench and so on, as part of the WISE to explore unobservable scientific phenomena. The inquiry activity helps students to connect new concept to their existing concept. The knowledge integration pattern was implemented in synthesize of WISE. This pattern involves four main learning processes. (i) Eliciting student concept, consistent with the values of predictions; (ii) adding new concept, essential case or new scientific concepts; (iii) supporting students to distinguish among the new and existing concept by constructing arguments and using scientific evidence; and (iv) fostering reflection and self-monitoring to help students consolidate their understanding. Table 3 shows an example of inquiry activity based on knowledge integration (KI) processes. The researchers will assess students’ conceptual understanding and knowledge integration by analyzing their answers the prompting questions embedded in WISE activity before and after investigation.

Table 3: Embedded prompts/assessment capturing the knowledge integration processes.

<table>
<thead>
<tr>
<th>KI processes</th>
<th>Description and Examples of activity</th>
<th>Illustrations of WISE screenshot</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Eliciting ideas</td>
<td>Using prompting questions to probe prior knowledge and existing ideas</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Example:</td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Could you draw the model of arrangement of particle in each state, liquid solid and gas?</td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Explain why did you draw the previous model.</td>
<td></td>
</tr>
<tr>
<td>2. Adding new ideas</td>
<td>Providing necessary information to help students make sense of the topic and connect to existing ideas</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Example:</td>
<td></td>
</tr>
<tr>
<td></td>
<td>- The picture below represents the arrangement of particle in solid, liquid and gas state of matters, you can move the cursor anywhere inside the picture to see the difference of particle movement.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>- For this part, you can control the molecular visualization to explore the structure of a solid, liquid and gas at the molecular level.</td>
<td></td>
</tr>
</tbody>
</table>
3. Distinguishing ideas

Encouraging students to distinguish among ideas and realizing how existing ideas relate to, conflict with, or extend these new ideas

Example:
- By the evident from previous page, recall the structure of a solid, liquid and gas at the molecular level in the model you investigated. Could you match the motion of particle with solid, liquid or gas?

4. Sorting out ideas

Encouraging students to sort out and refine their knowledge based on these evaluations

Example:
- Recall the previous questions. After your investigation, could you draw the model of arrangement of particle in each state, liquid solid and gas? And explain why did you draw the previous model.

6. Conclusion and Future work

The blended learning environment designed based on misconceptions of state of matter and phase change finding in this pilot study will be used in eleventh-grade students to evaluate the effectiveness of the proposed approach. In further study, the combination of CBILD and WISE will be used to enhance eleventh-grade students’ conceptual understanding, knowledge integration of the state of matter and phase change, and also scientific motivation. The simultaneous mixed methods strategy of non-equivalent control group design and phenomenological research design will be carried out. The participants will be separated into two group, control and experiment group. Only ILD will be implemented in a control group, on the other hand, ILD combined with WISE will be implemented in an experiment group. Student’s knowledge integration will be addressed during learning through WISE, which is designed based on knowledge integration framework, by taking the questions before and after using WISE within class period. To compare difference the outcome between two groups, pre-test, post-test and embedded questions in WISE will be analyzed and interpreted to answer the research question; does blended learning combining CBILD and WISE effect eleventh-grade students’ conceptual understanding, knowledge integration about state of matter and phase change, and scientific motivation differently from using only CBILD?

Acknowledgments

This work was financially supported by the Institute for the Promotion of Teaching Science and Technology (IPST) and Graduate School, Khon Kaen University. The author would like to express gratefully acknowledge to Faculty of Education, Khon Kean University, for supporting this contribution.
References


Investigating Correlation between Attitude toward Chemistry and Motivation within Educational Digital Game-based learning

Nattida NANTAKAEW\textsuperscript{a} & Niwat SRISAWASDI\textsuperscript{b*}
\textsuperscript{a}Science and Technology Education Program, Faculty of Education, Khon Kaen University, Khon Kaen, Thailand
\textsuperscript{b}Division of Science, Mathematics, and Technology Education, Faculty of Education, Khon Kaen University, Khon Kaen, Thailand
*niwsri@kku.ac.th

Abstract: Educational digital game-based learning, which is one of emerging pedagogies, could promote students’ motivation. With the benefit of the educational digital game-based learning, in this study, the educational digital game was designed for learning chemistry. To investigate the effect of attitude toward chemistry lesson on students’ motivation when they learn by using educational digital game, the study was conducted on eleventh grade students. The results show that attitude of interest in chemistry lesson makes students understanding and learning chemistry. The importance of chemistry in real-life, chemistry and occupational choice were not related to students’ motivation in attention, relevance, confident and satisfaction. Moreover, the students have a positive effect motivation after playing the educational digital game. To this end, this study concludes that the use of educational digital game-based learning could support students’ motivation even they have a positive or negative attitude in learning.

Keywords: Educational digital game, attitude toward chemistry, motivation

1. Introduction

Chemistry, which is the one of most important discipline, explains daily life phenomena. Chemistry concept related to other concepts in science such as the biology, physics and materials science. The nature of chemistry is abstract content which need to use imagination for connecting to real life situation. Chemistry requires three different levels of representation which are macroscopic, submicroscopic, and symbolic level. The topic of adhesive force in chemistry uses three levels of representation for explaining the phenomena. This topic related to understanding the basic phenomena in the science curriculum that student is incomprehensible (Eilam, 2004; Leite, Mendoza and Borsese, 2007). There are many factors related to difficulty in learning chemistry. Sirhan (2007) indicated the main factor of the learning difficulty in chemistry, curriculum content, and overload of students’ working memory space, motivation, language and communication. Therefore, motivation and attitude to learn is challenging to study, because they are major factors that effect on the success of learning, achievement and willingness of students. The students do not have motivation to understand if they perceive that those contents are difficult. In other words, format of instruction affects students’ attitude toward learning. If teacher could provide student-center instruction, students might be willing to learn.

In contemporary education, proper education needs to focus on students. Instruction based on interests, needs and abilities of students will persuade students willing to learn more, motivate them to learn led to get meaningful learning. Students in digital age interest in implementing technology. So, computer and communication technology, which is used widely today, can involve in education for supporting students’ learning. It also provides opportunity for the students to understanding of both basic and in-depth content. Moreover, it may make them to understand the complex process and can apply knowledge into everyday life. Learning by using the educational
digital game-based learning is a model that combines computer technology into the curriculum, so that the students gain both fun and knowledge (McNamara, Jackson and Graesser, 2010). So, the main challenge for educational research is combining digital game and instruction features in chemistry class. San Chee and Tan (2012) designed and develop an educational game named Legends of Alkhimia. They found that the game effectively fosters learning and supports conceptual understanding of chemistry. Moreover, Papastergiou (2009) found that students who learn via the computer game have more motivational than the non-gaming approach. So, educational computer games can be exploited as effective and motivational learning environments. The previous research has indicated that educational digital game in classroom can support learning and increase students’ motivation. However, chemistry class in Thailand lack of combining game into the classroom. Before including game with appropriate chemistry instruction, the attitude toward chemistry should be concerned. Therefore, this study investigates correlation between attitude toward chemistry and motivation in learning via playing game.

2. Literature Review

2.1 Educational Digital Game

The new media and digital technology industries and digital gaming immerse several environments. Digital games consist of dazzling and sophisticated images and sounds, alongside textual communication. Players get engagement which is both pleasurable and challenging. The educational digital game keep players immersed in digital worlds, knowledge, information, and skill development become increasingly accessible outside confines of formal education (Castell, Jenson and Taylor, 2007). Currently, educators employed digital game that insert content of subject matter or information for educational purpose. Several research presented empirical evidences that the educational digital games have positive effect on student learning. It improved not only learning achievement but also learning attitude and motivation to learn (Giannakos, 2013; Pilli and Aksu, 2013; Sung and Hwang, 2013).

2.2 Game-based Learning

In the past, game produce only for entertainment but recently educational researchers have attempted to adapt games for learning which calls educational games or serious game (Sorensen and Meyer 2007; Stone 2008). The games that compose of challenge, control, curiosity and fantasy can motivate persistence and enjoyment (Toro-Troconis and Partridge, 2010). The educators have developed games for three goals including: (i) students can learn from playing the game; (ii) the component of game can support learning; and (iii) students have motivation to learn when they learning by playing the game (McNamara, Jackson, & Graesser, 2010). Game-based learning is a kind of constructivist-based active learning. Based on the learning research, Watson, Mong and Harris (2011) found that using game in classroom made a shift of teaching from teacher-centered learning environment to student-centered learning environment.

2.3 Attitude and Motivation

The most important students’ characteristic associated with successful studies is attitude, motivation, and genuine interest (Berg, 2005). Attitudes and motivation are both important factors for the learning process. Success in learning, positive attitudes to learning and motivation to learn are linked. The two major factors influencing attitudes towards a subject are teacher quality and curriculum quality. They strongly influenced by the perceived curriculum relevance, in the sense that the learner perceives what is taught being related to their lifestyle (Sirhan, 2007). Moreover, Hofstein and Mamlok-Naaman (2011) suggested the three key factors that should be considered for enhancing attitudes and interests are the methods used to present the content, instructional techniques, and gender issues.
3. Purpose

The goal of this study was to investigate correlation of attitude toward chemistry with motivation in learning after playing the educational digital game in the topic of adhesive force and to explore students’ motivation after providing the game. Specifically, the research questions were answered:
- How were the influences of attitude toward chemistry on the students’ motivation after providing the educational digital game?
- Is it suitable to implement the educational digital game in a Thai school?

4. Methods

4.1 Study Participants

This study recruited 37 students who are studying in eleventh grade, age ranging from 17 to 18 years in a local public school at the northeastern region of Thailand. They are enrolling program that emphasizes using science and technology in the classroom. They also studied about adhesive force which be contained in the topic of properties of liquid in the last semester.

4.2 Instruments

This research used two instruments for determining students’ attitude towards the chemistry lesson and motivation in learning via the game. First, the attitude scale developed from Attitude Towards Chemistry Lesson Scale (ATCLS) of Ayyıldız and Tarhan (2013) consisting of 25 items. All items were classified into four scales, including interest in chemistry lessons (six items), understanding and learning chemistry (ten items), importance of chemistry in real-life (five items), and occupational choice related to chemistry (four items). Its Cronbach’s alpha reliability coefficient of this instrument was 0.88, implying that it is reliable. Each scale of ATCLS has Cronbach’s alpha reliability coefficient from 0.52 to 0.82. The sample item and description of each scale are provided in Table 1. Second, the motivation in learning via the game investigated by using Instructional materials motivational survey (IMMSS) developed from Huang (2011). This instrument consists of 18 items which are divided into attention (eleven items), relevance (four items), confident (three items) and satisfaction (one item). It Cronbach’s alpha reliability coefficient was 0.86. Each scale of ATCLS has Cronbach’s alpha reliability coefficient from 0.52 to 0.83.

<table>
<thead>
<tr>
<th>Scale</th>
<th>Description</th>
<th>Sample item</th>
</tr>
</thead>
<tbody>
<tr>
<td>Attention</td>
<td>Extent to which students’ response to perceive instructional stimuli provided by the game.</td>
<td>The game has things that stimulated my curiosity.</td>
</tr>
<tr>
<td>Relevance</td>
<td>Extent to which student connect their prior learning experience with the game.</td>
<td>There were examples that showed me how the game could be important to some people in the learning setting.</td>
</tr>
<tr>
<td>Confidence</td>
<td>Extent to which student has positive expectation after finishing learning activity.</td>
<td>The game had so much information that it was hard to pick out and remember the important points.</td>
</tr>
<tr>
<td>Satisfaction</td>
<td>After learning via the game, student accept practice newly acquired knowledge</td>
<td>It felt good to successfully complete the game.</td>
</tr>
</tbody>
</table>
Table 2 shows the sample item and description for four scales. The answers of students in both ATCLS and MLG scale were labeled as strongly agree (5 point), agree (4 points), partly agree (3 point), disagree (2 point) and strongly disagree (1 point) from the positive to the negative.

<table>
<thead>
<tr>
<th>Scale</th>
<th>Description</th>
<th>Sample item</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interest in chemistry lesson</td>
<td>Extent to which student preferred chemistry learning.</td>
<td>I would like the teaching period of the chemistry lesson more often.</td>
</tr>
<tr>
<td>Understanding and learning chemistry</td>
<td>Extent to which student developed themselves and implicated in chemistry easily.</td>
<td>I find using chemical symbols to be easy.</td>
</tr>
<tr>
<td>The importance of chemistry in real life</td>
<td>Extent to which student thought chemistry were appropriate to real-life.</td>
<td>I believe that chemical knowledge helps us interpret seriously events in our daily life.</td>
</tr>
<tr>
<td>Chemistry and occupational choice</td>
<td>Extent to which student use the information learned in the chemistry classroom for the futuristic work.</td>
<td>My career could be chemist/chemistry teacher/chemical engineer.</td>
</tr>
</tbody>
</table>

4.3 Learning Material

In this study, the design of the game, called “The Pipe”, was related to content of adhesive force. The game provides problem situation to students. The game engages students to imagine macroscopic level linked to real life phenomena and also provides information which are three level of representation in chemistry including macroscopic, sub-microscopic and symbolic level. The students were asked to use those information to solve problem. Figure 1 illustrates procedures of learning activity.

4.4 Data Collection and Analysis

The intervention class consists of 37 students. Before providing The Pipe game, students was surveyed the attitude toward chemistry lesson scale. The Instructional materials motivational survey was provided to the students after they interacted with the developed game. The data from two scales reflected the relation in each variable by using Pearson’s correlation in SPSS. The result of Pearson’s correlation describe relations of interest in chemistry lesson, understanding and learning chemistry, importance of chemistry in real-life, chemistry and occupational choice in attitude toward chemistry and each scale of motivation including attention, relevance, confident and satisfaction. The influence of attitude toward chemistry on learning motivation via playing the game was investigated.
(A) The game starts with problem situation in the factory. The problem is transferring water through pipes slowly. The player as chemist is asked to choose the proper pipe for the flow of water.

(B) In the game provides scaffolding for decision, the player can see molecular structure of water and each type of pipe. In this part, the player can observe experiment demonstration of the flow of water. However, the player needs to pay coins for seeing them.

(C) This is one sample of the molecules in scaffolding part.

(D) After students click to play a game, it goes in this part in which they pay coins for the various shapes of each pipe for connecting two fix pipes to each other. Students have to concern about budget that they have.

(E) After the player put the proper pipes, they need to click OK bottom for confirming decision.

(F) The last part of the game shows money left, the number of pipes and time used for flowing of water is shown.

Figure 1. Example of chemistry learning activity on properties of liquid.
5. Results and Discussion

5.1 Correlation between Attitude and Motivation

Table 3 shows Pearson’s correlation of Interest in Chemistry Lesson (ICL), Understanding and Learning Chemistry (ULC), Importance of Chemistry in Real-life (ICR), Chemistry and Occupational Choice (COC) in ATCLS and Attention (A), Relevance (R), Confident (C) and Satisfaction (S) in IMMSS. Mean and standard deviation are also presented in Table 3.

Table 3: Descriptive and correlation for attitude toward chemistry lesson and motivation.

<table>
<thead>
<tr>
<th>Scale</th>
<th>ICL</th>
<th>ULC</th>
<th>ICR</th>
<th>COC</th>
<th>A</th>
<th>R</th>
<th>C</th>
<th>S</th>
</tr>
</thead>
<tbody>
<tr>
<td>ICL</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ULC</td>
<td>0.57**</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ICR</td>
<td>0.21</td>
<td>0.36*</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>COC</td>
<td>0.66**</td>
<td>0.563*</td>
<td>0.61**</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>0.13</td>
<td>0.10</td>
<td>0.19</td>
<td>0.16</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R</td>
<td>0.18</td>
<td>0.04</td>
<td>0.06</td>
<td>0.11</td>
<td>0.78**</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>0.16</td>
<td>0.21</td>
<td>0.11</td>
<td>0.13</td>
<td>0.64**</td>
<td>0.61**</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>S</td>
<td>0.14</td>
<td>0.05</td>
<td>0.02</td>
<td>-0.06</td>
<td>0.67**</td>
<td>0.57**</td>
<td>0.74**</td>
<td>1</td>
</tr>
<tr>
<td>Mean</td>
<td>16.84</td>
<td>29.71</td>
<td>17.05</td>
<td>12.89</td>
<td>36.87</td>
<td>9.90</td>
<td>13.64</td>
<td>3.74</td>
</tr>
<tr>
<td>SD</td>
<td>3.19</td>
<td>3.178</td>
<td>2.86</td>
<td>2.70</td>
<td>7.11</td>
<td>1.81</td>
<td>2.60</td>
<td>1.03</td>
</tr>
</tbody>
</table>

**p < 0.01  
*p < 0.05

Regarding Pearson’s correlation analysis of each scale from ATCLS, interest in chemistry lesson was positively related to Understanding and Learning Chemistry (ULC), and chemistry and occupational choice. Understanding and Learning Chemistry (ULC) was positively related to importance of chemistry in real-life (ICR) and chemistry and occupational choice (COC). In addition, importance of chemistry in real-life (ICR) was positively related to chemistry and occupational choice (COC). All scale positively related together except interest in chemistry lesson (ICL) scale that do not relate to importance of chemistry in real-life (ICR). These results imply that students attend to chemistry lesson because they understand chemistry concept and want to work in career related chemistry.

From the result of IMMSS, attention (A), relevance (R), confident (C) and satisfaction (S) are linked together. From the findings, it suggests that if students have only one scale of attention (A), relevance (R), confident (C) and satisfaction (S), they have motivation to learn via games. Consider Table 3, the interest in chemistry lesson (ICL), understanding and learning chemistry (ULC), importance of chemistry in real-life (ICR), chemistry and occupational choice (COC) was no related to attention (A), relevance (R), confident (C) and satisfaction (S) when provided the Pipe game to students. So, the educational digital game can use for all students even if they have a negative or positive attitude toward chemistry.

The findings from previous study revealed that the educational digital game improve both attitude toward learning and motivation, because it furnished more attractive of learning environment (Eseryel, Law, Ifenthaler, Ge and Miller, 2014; Sung and Hwang, 2013). But we do not know about the effect of attitude on motivation during interaction with the educational digital game. This study indicated that motivation in learning via game does not depend on attitude toward chemistry. Although students negative or positive attitude toward chemistry, they can learn chemistry by playing game.
5.2 Students' Responses

The scale of motivation in IMMSS was summarized in Table 4 which shows the means and standard deviation of each scale. We found that students had high level of motivation in attention, relevance, confident and satisfaction after participating in the developed game.

Table 4: Scale means and summary response.

<table>
<thead>
<tr>
<th>Scale</th>
<th>Mean (SD)</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Attention</td>
<td>36.87(7.11)</td>
<td>High level</td>
</tr>
<tr>
<td>Relevance</td>
<td>9.90 (1.81)</td>
<td>High level</td>
</tr>
<tr>
<td>Confident</td>
<td>13.64 (2.60)</td>
<td>High level</td>
</tr>
<tr>
<td>Satisfaction</td>
<td>3.74 (1.03)</td>
<td>High level</td>
</tr>
</tbody>
</table>

Many previous studies presented that learning via playing game support perspective of cognition and affection. From evidence of achievement test, the game can develop students’ performance. Students understand content easily and deeply. For affective domain, the game increase students’ motivation. Students preferred to learn by playing game, because they had a positive perspective in learning. They also perceived that the game is useful, easy, interested and enjoyable. All of them can motivate students to learn. We assure that game support learning effectively (Cheng, Huang, and Chen, 2013; Sung and Hwang, 2013).

6. Conclusion and implementation

The result of this study provided a more understand on influences of attitude toward subject matter, for example, chemistry, on student's motivation to learn in a setting of educational game learning environment. The finding indicated that attitude toward chemistry have no correlation on motivation to learn chemistry via game. Thus, we can use the educational digital game for participants who have both a positive and negative effect. Although they like or dislike to learn chemistry lesson, they remain have a positive effect after learning via educational digital game.

Although there are many researches indicated that teaching and learning via game improve students’ motivation, we should collect pre- and post-motivation for comparing motivation before and after learning. In an addition, the challenge is how to immerse the digital game into classroom instruction. A previous study by San Chee and Tan (2012) used educational game to support students’ inquiry learning process, and they found that the students can effectively inquire to learn science through digital game. Based on the findings of this study, we will design educational digital game about properties of liquid use Student-Associated Game-based Inquiry (SAGOI) approach for improving chemistry learning in quasi-experimental design that include two different-intervention groups of students. One group will provide SAGOI instruction and another acquire traditional instruction. The mixed research methodology combined quantitative method of non-equivalent control group design with qualitative method of phenomenological research design will carry out in future research.

Acknowledgements

This work was financially supported by the Institute for the Promotion of Teaching Science and Technology (IPST) and Graduate School, Khon Kaen University. The author would like to express gratefully acknowledge to Faculty of Education, Khon Kean University, for supporting this contribution.
References


Development and preliminary evaluation of a knowledge management-based online teacher community platform for science fair instruction: A cluster analysis

Chiu-Ming, Hu, Chao-Shen Cheng, Li-Jen Wang, Huei-Tse Hou, Yi-Chun Kuo, Cheng Teng Yao & Ying-Tien Wu*

*Graduate Institute of Network Learning Technology, National Central University, Taiwan
bGraduate Institute of Applied Science and Technology, National Taiwan University of Science and Technology, Taiwan
cCentral Taiwan Teaching/Learning Resource Center, Feng Chia University, Taiwan
*ytwu@cl.ncu.edu.tw

Abstract: Inquiry is the core of modern science education. In science classes, conducting science fair projects is one of the most common inquiry activities. Through conducting these projects, learners have the chance to carry out open inquiry which may help them build deeper understanding of science knowledge, concepts, science skills, and positive attitudes toward science. However, previous research has revealed that teachers encountered many challenges when implementing inquiry-based instruction, such as insufficient time, sources, professional knowledge and experience of inquiry learning. To address the important issue, by integrating community-based knowledge management (KM) tools and personal-based knowledge management (PKM) tools, a KM-based online teacher community platform was developed and preliminarily evaluated in this study. Moreover, cluster analysis was also conducted to categorize the participants’ attitude-behavior patterns into different clusters, and the attitudes and behaviors of these different clusters were further analyzed. A total of 103 volunteer Taiwanese elementary school science teachers participated the system evaluations of the Teacher Science Fair Instruction Knowledge Management System (TSFI-KMS) in this study. Their responses on the quantitative questionnaire designed for the system evaluation in this study showed that they expressed satisfactory perceived usefulness and ease of use of the TSFI-KMS. Also, they expressed high willingness to use the TSFI-KMS for professional development regarding inquiry-based instruction. Further analysis indicate that the teachers’ perception and usage preference toward TSFI-KMS might be influenced by their academic background and experience of using social media. In this study, suggestions on teaching practices, improvement on the system design, and future research are also discussed.

Keywords: Inquiry; science fair; knowledge management; cluster analysis

1. Introduction

Inquiry is the core of modern science education. In science classes, conducting science fair projects is one of the most common inquiry activities. Through conducting these projects, learners have the chance to carry out open inquiry which may help them build deeper understanding of science knowledge, concepts, science skills, and positive attitudes toward science. However, the literature revealed that many teachers may lack professional knowledge, time, resources, and assistance when conducting science fair instruction (Anderson, 2002), and only a few science teachers know how to guide students to conduct such projects or inquiry activities effectively (Justi & Gilbert, 2002). Therefore, supporting science teachers to develop their pedagogical content knowledge (PCK) (i.e., professional knowledge) regarding science fair instruction, as well as providing them with resources and assistance for science fair instruction, is crucial.
To address the important issue mentioned above, online communities, which have been advocated as a potential tool for teachers to promote their professional development, would be helpful. However, an online platform for teacher communities focusing on science fair instruction is still not yet available. Therefore, this study aimed to develop such a platform. Moreover, the management of professional knowledge created by an online community is always an important issue for both the online community and its members. However, most online teacher community platforms are not formulated based on knowledge management perspectives. By integrating community-based knowledge management (KM) tools (Spector, 2002) and personal-based knowledge management (PKM) tools (Tsui, 2002), a KM-based online teacher community platform was developed and preliminarily evaluated in this study. Moreover, cluster analysis (Hou et al., 2011) was also conducted to categorize the participants’ attitude-behavior patterns into different clusters, and the attitudes and behaviors of these different clusters were further analyzed.

2. System development

In this study, the Teacher Science Fair Instruction Knowledge Management System (TSFI-KMS) was developed based on the KM and PKM theoretical framework proposed by Spector (2002) and Tsui (2002). TSFI-KMN is an online knowledge management environment. To meet different teachers’ preferences for using the KM system, there are two different portals in TSFI-KMS: the “personal-based portal” and the “community-based portal.”

The system framework of TSFI-KMS consists of several main KM and PKM modules and a knowledge base which stores members’ profiles, knowledge sharing process and knowledge documents. Examples of detailed functions of the modules are shown in Table 1.

<table>
<thead>
<tr>
<th>Portal</th>
<th>Knowledge management tool phase</th>
<th>Module</th>
<th>Function examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>KM-based (community-based portal)</td>
<td>Communication</td>
<td>Communication module</td>
<td>Asynchronous communication tools, project progress reports</td>
</tr>
<tr>
<td></td>
<td>Coordination</td>
<td>Coordination module</td>
<td>Project calendar, coordination tools</td>
</tr>
<tr>
<td></td>
<td>Collaboration</td>
<td>Collaboration module</td>
<td>Images, videos, and document sharing, science fair project collaboration</td>
</tr>
<tr>
<td>PKM-based (personal-based portal)</td>
<td>Indexing and information capturing/management</td>
<td>Searching and Information capturing module</td>
<td>Information indexing, searching results combination, information capturing, information alert, information/documents uploading</td>
</tr>
<tr>
<td></td>
<td>Personal communication management/analysis</td>
<td>Communication management module</td>
<td>E-mail management, communication message analysis</td>
</tr>
<tr>
<td></td>
<td>Learning profile management</td>
<td>Personal profile management module</td>
<td>Learning process tracking, learning profile uploading</td>
</tr>
</tbody>
</table>
3. Methodology (System evaluation)

3.1 Participants

The participants of this study were 103 volunteer Taiwanese elementary school science teachers, including 41 male teachers and 62 female teachers. Regarding the teaching experience distribution of the participant teachers, 23 teachers (22.3%) had less than 5 years, 40 (38.9%) had between 5 and 9 years, 24 (23.3%) had between 10 and 14 years, and 16 (15.5%) had over 15 years. They also had various experience of science fair instruction and of using online social media, such as forums, Facebook, and blogs.

3.2 Evaluation procedures

When the participant teachers first logged into the TSFI-KMS platform, they were given a brief introduction to how to use the system. Then, the participant teachers’ background information was collected through an online questionnaire before they started to explore the platform. The participants were asked to explore TSFI-KMS by themselves whenever they had free time during a period of two weeks. After the exploration task, the teachers were asked to evaluate TSFI-KMS by completing online questionnaires.

3.3 Instruments and data collection

In this study, an integrated system evaluation involving both attitude and behavior perspectives was conducted. With online questionnaires, this study collected the data regarding attitude evaluations:

1. Teachers’ acceptance of TSFI-KMS:
   The 6-point Likert-scale questionnaire developed in Phang et al. (2009) was modified and used in this study. The modified instrument consists of two scales: usefulness (6 items) and usability (7 items). In this study, the overall alpha reliability value of the instrument is 0.93.

2. Teachers’ perceived usefulness of the KM and PKM tools in TSFI-KMS:
   Two 6-point Likert-scale instruments consisting of 12 and 18 items for assessing teachers’ perceived usefulness of the KM and PKM tools in TSFI-KMS were developed in this study. The alpha reliability values of the two instruments are 0.92 and 0.94, respectively.

3. Teachers’ perceived usefulness of TSFI-KMS for improving science fair instruction PCK:
   A 6-point Likert-scale instrument with 5 items developed in this study was used to evaluate the teachers’ perceived usefulness of TSFI-KMS for improving their science fair instruction PCK.

The current study also collected data regarding the teachers’ participation in evaluating TSFI-KMS. The teachers’ participation in evaluating the two portals of TSFI-KMS was evaluated respectively by counting their number of clicks when evaluating the two portals of the system.

4. Major findings and Conclusions

4.1 Teachers’ attitudes toward TSFI-KMS and participation in evaluating TSFI-KMS

Table 2 shows that the participating teachers expressed high acceptance of TSFI-KMS (mean=5.03). Also, they highly recognized the usefulness of the KM tools (mean=5.22) and PKM tools (mean=5.13) provided by TSFI-KMS, and agreed with the usefulness of TSFI-KMS for improving their science fair instruction PCK (mean=5.19). Moreover, the teachers had relatively more participation in the community-based portal evaluation (mean=14.74) than in the personal-based portal evaluation (n=6.41), indicating that the teachers in this study might be more oriented towards using the community-based portal of TSFI-KMS.
Table 2: Teachers’ attitudes toward TSFI-KMS and participation in platform evaluation

<table>
<thead>
<tr>
<th>Attitudes toward the platform</th>
<th>mean</th>
<th>S.D.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acceptance of TSFI-KMS</td>
<td>5.03</td>
<td>0.59</td>
</tr>
<tr>
<td>Perceived usefulness of the KM tools</td>
<td>5.22</td>
<td>0.49</td>
</tr>
<tr>
<td>Perceived usefulness of the PKM tools</td>
<td>5.13</td>
<td>0.44</td>
</tr>
<tr>
<td>Perceived usefulness for improving PCK</td>
<td>5.19</td>
<td>0.59</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Participation in platform evaluation</th>
<th>mean</th>
<th>S.D.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Community-based portal evaluation</td>
<td>14.74</td>
<td>10.43</td>
</tr>
<tr>
<td>Personal-based portal evaluation</td>
<td>6.41</td>
<td>6.06</td>
</tr>
</tbody>
</table>

4.2 Cluster analysis of the participants’ attitude-behavior patterns

A cluster analysis using the attitude and behavior indicators (as revealed in Table 1) was further conducted in this study. We first conducted a hierarchical cluster analysis, then a Ward method’s dendrogram to determine the appropriate number of clusters. The teachers’ attitude-behavior patterns are divided into three clusters, as shown in Table 3.

According to Table 3, among the three teacher groups, the teachers in group B (about 3.9%) most actively participated in the community-based portal evaluation, while they had the lowest participation in the personal-based portal evaluation. Moreover, these teachers also expressed the most positive attitudes toward TSFI-KMS and the KM and PKM tools, indicating that they strongly recognized the usefulness of the KM and PKM tools. Consequently, they are more likely to continue to use TSFI-KMS, with an orientation towards using the community-based portal. These teachers with higher acceptance of the KM tools may more actively participate in an online community, and thus are likely to play significant roles in the social interactions within the online community. In other words, these teachers are usually the leaders of knowledge sharing or the mediums of knowledge exchange within online communities. Therefore, meeting these teachers’ needs in improving their PCK is crucial for promoting the depth of knowledge sharing within TSFI-KMS. To this end, refinements to TSFI-KMS should be made in future research.

Table 3: The result of the cluster analysis

<table>
<thead>
<tr>
<th>Attitude and behavior indicators</th>
<th>Clusters</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>A</td>
</tr>
<tr>
<td>Acceptance of TSFI-KMS</td>
<td>(n=64,  62.1%)</td>
</tr>
<tr>
<td>Perceived usefulness of the KM tools</td>
<td>5.01</td>
</tr>
<tr>
<td>Perceived usefulness of the PKM tools</td>
<td>5.14</td>
</tr>
<tr>
<td>Perceived usefulness for improving PCK</td>
<td>5.10</td>
</tr>
<tr>
<td>Community-based portal evaluation</td>
<td>9.14</td>
</tr>
<tr>
<td>Personal-based portal evaluation</td>
<td>5.38</td>
</tr>
</tbody>
</table>

The teachers in group C (34%) most keenly participated in the personal-based portal evaluation, and also expressed the most positive perceptions of the usefulness of TSFI-KMS for improving their science fair instruction PCK. It seems that these teachers recognize the usefulness of TSFI-KMS for improving their PCK. Thus, they might also be willing to use TSFI-KMS in the future, with an orientation towards using the personal-based portal. Moreover, compared with most of the teachers in this study (i.e., Cluster A), they perceived higher usefulness and technology
acceptance based on TSFI-KMS. However, they had relatively lower participation in the community-based portal, which might restrain the interaction and knowledge sharing with other teachers. Therefore, how to integrate the community-based and personal-based portals more effectively should also be an important issue in the refinement of TSFI-KMS in future studies.

Although the teachers in group A (62.1%), in general, had relatively lower average scores on the attitude indicators compared with the other two groups, their average scores for these indicators were still greater than 5, revealing that they still expressed positive attitudes toward TSFI-KMS. Nevertheless, they had relatively lower average scores for participation in the platform evaluation (i.e., the mean score of the community-based portal evaluation plus the mean score of the personal-based portal evaluation), indicating that they might not be as eager as the teachers in the other two groups to use TSFI-KMS. Besides, compared with the personal-based portal evaluation, the teachers in group A were likely to be more oriented towards using the community-based portal. Since the main objective of TSFI-KMS is to improve teachers’ professional development in PCK, suitable scaffolding tools provided by the platform for motivating science teachers’ online professional development should be addressed in future research.

In conclusion, the teachers in this study expressed high satisfaction with and acceptance of TSFI-KMS, and scored highly on the usefulness of the KM and PKM tools in TSFI-KMS, and the usefulness of TSFI-KMS for improving science fair instruction PCK. Therefore, the TSFI-KMS platform is suitable for teachers’ professional development in science fair instruction. Besides, the cluster analysis in this study revealed that the teachers in the different clusters had their own preferences and attitudes toward the use of KM tools, providing implications for platform refinements, the formation and the management of online teacher communities, and teacher continuing professional development.

Acknowledgment
This study was funded by the National Science Council, Taiwan, ROC, under grant contract number NSC 101-2628-S-008-001-MY3, but the opinions expressed in this article do not reflect the position of the National Science Council.

References
An Evaluation of Macro-Micro Representation-based Computer Simulation for Physics Learning in Liquid Pressure: Results on Students’ Perceptions and Attitude

Jarunya BUYAIa, Niwat SRISAWASDIb*

aScience and Technology Education Program, Faculty of Education, Khon Kaen University, Khon Kaen, Thailand
bDivision of Science, Mathematics, and Technology Education, Faculty of Education, Khon Kaen University, Khon Kaen, Thailand
*niwsri@kku.ac.th

Abstract: Computer simulation has been widely used to enhance teaching and learning for last decades and researchers mentioned that the use of computer-simulated experimentation can actively engage and enhance student’s meaningful learning in subject contents. As such, a computer simulation for physics learning in liquid pressure has been created with regarding the interplay among macro- and micro representation of physical knowledge. To evaluate the developed simulation, 40 twelve-grade students were recruited to participate with the simulation. A Likert-scale perception and attitude questionnaires were administered to the students before the participation as pre-test. The perception questionnaire was, only, distributed to the student as post-test. The repeated-measures MANOVA results indicated that there was no significant main effect on gender and interaction effect between gender and time (pre-test/post-test), but there were significant main effect on time. That is, there was no effect of gender difference on students' perceptions towards physics learning through computer simulation. Moreover, Pearson’s correlation indicated that the computer simulation could be used to promote physics learning experience for all students even if they have a negative or positive attitude toward physics lessons

Keywords: Computer simulation, interactive experience, physics learning, content representation

1. Introduction

Currently, innovative technologies in science teaching and learning is growing continuously. The use of learning technologies such as computer simulation offers students an interactive learning experience and allows students to learn on their own way (Vreman-de Olde et al. 2013). Computer simulations have become increasingly powerful and available to teachers in the past decade (Trundle and Bell, 2010). In community of science and technology education, computer simulation is recognized as a pedagogical tool to support conceptualization of science concepts and facilitate process of scientific inquiry or discovery by visualizing and interacting with dynamics models of natural phenomenon (Ton and Joolingen, 2012; Perkins et al., 2006; Wieman, Perkins and Adams, 2008). These technology offer idealized, dynamic and visual representations of invisible phenomena and experiments which would be dangerous, costly or otherwise not possible in school laboratories (Hennessy, 2007). In addition, researchers found that a learning environment with computer simulation has the advantages that students can
systematically explore hypothetical situations, in a realistic environment without stress, in comparison with textbooks and lectures (Rutten, Joolingen, and Veen, 2012). In physics education, pressure of liquid is a fundamental concept for student’s comprehension about fluid mechanics concepts. A few studies have reported that students often encounter learning difficulties and hold unscientific understanding of this concept. There were many misconceptions about the liquid pressure among high school students (Kariotoglou and Psillos, 1993). For example, most of students believed that the shape of the container and amount of the liquid is effective on liquid pressure (Sahina, Çepni, and Ipek, 2010). According to this problem, this study aims to develop a macro-micro representation-based computer simulation for physics learning in liquid pressure to facilitate student's conceptual comprehension of these physics concepts. In this study, a preliminary findings on secondary school students' perceptions was reported.

2. Application of Computer Simulation in Science Education.

Computer simulations are a powerful instructional tools which has been recognized by the community of science education by presenting theoretical, experimentation, models of real-world components, phenomena, processes in science, for students in order to observe, explore, recreate, and receive immediate feedback about phenomena and processes. To support effective science learning activities, computer simulation was used to facilitate their learning difficulties and abstract and complicated content as conceptual learning tools. (Chen et al., 2011; Colella, 2000; de Jong and Van Joolingen, 1998). In an addition, computer simulation helps learners to understand chemical or biological phenomena which are not able to observe directly (Cook 2006; Wu and Shah 2004). To create awareness through higher-order thinking skill, function of real-time data displays related to a dynamic phenomenon and information on how change parameters synchronously were employed into simulation-based teaching and learning. (de Jong and van Joolingen, 1998; Ronen and Eliahu, 2000). Even student hold alternative conceptions about science-related phenomena, computer simulation could be used to support a more meaningful learning in science concepts through the process of conceptual change (Srisawasdi and Kroothkeaw, 2014). Computer simulation has significant potential as a supplementary tool for effective conceptual change learning based on the integration of technology and appropriate instructional strategies. There are several educational values that computer simulation adds into science learning activities (Hennessy et al. 2006). As such, successful concepts of simulation-based teaching and learning have been reported by means of discovery learning (de Jong and van Joolingen, 1998; de Jong, Linn, and Zacharia, 2013) and inquiry-based learning (Perkins et al., 2006; Srisawasdi and Kroothkeaw, 2014; Wieman, Perkins, and Adams, 2008). With the importance, computer simulations are effectively linked to pedagogy as well. (Flick and Bell 2000). Following these class types, student can discover the principles, rules, and characteristics of scientific phenomena through change variable values and observe effects to form scientific conclusions in computer simulation (Veemans et al., 2006). Therefore, the use of computer simulation with these pedagogical approaches could be as instructional line for teaching and learning in school science.

3. Purpose

The goals of this study were to investigate students’ perceptions towards computer-simulated physics experimentation and correlation between attitude toward physics and physics motivation after interacting with the computer-simulated physics experimentation. Specifically, the following questions were answered:

1) Do the students engaged in computer-simulated physics experimentation perform significantly better by perceive learning, flow of learning experience, enjoyment, perceive ease of use, perceive of satisfaction, and perceive of usefulness?
(2) How were the influence of physics attitude on students’ perceptions after interacting with computer-simulated physics experimentation?

4. Method

4.1 Participants

A total of 40 twelfth grade students (female = 25, male = 15), age ranging from 17 to 18 years, in a local public school at the northeastern region of Thailand participated in this study. They were attending a physics course for basic education level. Regarding to prior learning experience, they have no experience yet using computer-simulated experiment in physics. This implied that they are heterogeneous on perception towards computer-simulated physics experimentation before interacting with the present experimental study.

4.2 Instructional Materials

The design of computer-simulated experimentation in physics of liquid pressure provides a rich context of representations where macro- and micro-scale representations were employed coordinately to visualize how the physical phenomena works. In additions, the computer-simulated physics experiment on liquid pressure has been produced by examining the attributes of the physics concept to provide information in which essential mental sets are needed to construct a scientific view of the concepts. According to this step, a concept map associated to learning liquid pressure simulation was constructed in hierarchical order of attributes. After the attributes and essential underlying concepts were identified by the researchers and two experts. As such, it involved three essential parameters (i.e. depth of water, shape of container, and density of liquid) which related to liquid pressure phenomenon, as shows in Figure 1.

![Figure 1. Example of screen interface of computer-simulated physics experiment on liquid pressure](image)

For the parameter of depth of water, the computer-simulated physics experiment on liquid pressure prepares to build student's conceptual knowledge on relationship between the depth of water and pressure, and serves physical understanding on what would happen to pressure if height of water is changed. In this part, student could interact with controlling levels of liquid and then measure its pressure by using a pressure gauge. Moreover, they could also see whether the liquid molecules behave relating to the depth. Figure 2 displays an example of interactive features in the computer-simulated physics experiment on relationship between liquid pressure and depth.
To facilitate student's conceptual knowledge on impact of liquid density on its pressure, the computer-simulated physics experiment on liquid pressure provides opportunity to select different types of liquid (i.e. ethanol, benzene, water, acetic acid, and glycerin) for investigating the concept. In this part, student could interact with selecting types of liquid and then measure its pressure. Moreover, they could also see whether the liquid molecules behave relating to the type of liquid. Figure 3 displays an example of interactive features in the computer-simulated physics experiment on relationship between density of liquid and its pressure.

Visualizing an effect of shape of container on liquid pressure was an aim for creating this computer-simulated experimentation. Figure 4 displays an example of interactive features in the computer-simulated physics experiment on relationship between shape of container and liquid pressure. Student could interact with the experiment by selecting a shape of container and then measure its pressure by using a pressure gauge. This part prepares to build student's conceptual knowledge on what would happen to pressure if shape of container is changed.
4.3 Instrument

A 21-item Likert-scale questionnaire was developed to use in this study for examining students’ perceptions towards the computer-simulated physics experiment on liquid pressure on six subscales: flow, enjoyment, perceived learning, perceive ease of use, perceive of usefulness, perceive of satisfaction. All of these 5-point Likert scale items obtained from (Cheng, 2014) and Barzilai and Blau (2014). From the English version, an identical version in Thai was constructed, and one expert was recruited to identify communication validity of the items. The respondents were required to consider each possible reason for computer simulation and rate how much the respondent agree with into five scale (1-strongly disagree; 2-disagree; 3-neutral; 4-agree; 5-strongly agree). The reliability for the overall questionnaire was 0.88 and for each subscale was presented in Table 1.

Table 1: Example items of perception questionnaire for each construct.

<table>
<thead>
<tr>
<th>Dimension</th>
<th>Sample items</th>
<th>α</th>
</tr>
</thead>
<tbody>
<tr>
<td>Perceive learning</td>
<td>▪ The simulation added to my knowledge.</td>
<td>0.803</td>
</tr>
<tr>
<td></td>
<td>▪ I learned new things from the simulation.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>▪ The simulation will help me remember the things I learned.</td>
<td></td>
</tr>
<tr>
<td>Flow</td>
<td>▪ I lost track of time when I played.</td>
<td>0.822</td>
</tr>
<tr>
<td></td>
<td>▪ I really got into the simulation.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>▪ I was very involved in the simulation.</td>
<td></td>
</tr>
<tr>
<td>Enjoyment</td>
<td>▪ I enjoyed the simulation.</td>
<td>0.745</td>
</tr>
<tr>
<td></td>
<td>▪ I had fun playing the game.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>▪ Playing the simulation was pleasant.</td>
<td></td>
</tr>
<tr>
<td>Perceive ease of use</td>
<td>▪ It is easy for me to learn how to use simulation.</td>
<td>0.737</td>
</tr>
<tr>
<td></td>
<td>▪ The user interface of simulation is easy to use.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>▪ I can easily accomplish what I need to do in simulation.</td>
<td></td>
</tr>
<tr>
<td>Perceive of usefulness</td>
<td>▪ Simulation can help me learn more effectively.</td>
<td>0.842</td>
</tr>
<tr>
<td></td>
<td>▪ Simulation can improve my course performance.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>▪ It is useful to study the course content with simulation.</td>
<td></td>
</tr>
<tr>
<td>Perceive of satisfaction</td>
<td>▪ I feel comfortable to use simulation.</td>
<td>0.774</td>
</tr>
<tr>
<td></td>
<td>▪ I enjoy the experience of using simulation.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>▪ I am willing to continue using simulation for learning in other courses</td>
<td></td>
</tr>
</tbody>
</table>

In addition, a 25-item Likert-scale questionnaire obtained from Ayyıldız and Tarhan (2013) was applied to measure attitude towards physics lessons. All items were classified into
four scales, including interest in chemistry lessons (6 items), understanding and learning physics (10 items), importance of physics in real-life (5 items), and occupational choice related to physics (4 items). Its cronbach’s alpha reliability coefficient of this instrument was 0.88, implying that it is reliable. Each scale has cronbach’s alpha reliability coefficient from 0.52 to 0.82. Table 2 shows the sample item and description for four scales.

Table 2: Scale descriptions and sample items for the physics attitude questionnaire

<table>
<thead>
<tr>
<th>Scale</th>
<th>Description</th>
<th>Sample item</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interest in physics lesson</td>
<td>Extent to which student preferred physics learning.</td>
<td>I would like the teaching period of the physics lesson more often.</td>
</tr>
<tr>
<td>Understanding and learning physics</td>
<td>Extent to which student developed themselves and implicated in physics easily.</td>
<td>I find using symbols in physics to be easy.</td>
</tr>
<tr>
<td>The importance of physics in real life</td>
<td>Extent to which student thought physics were appropriate to real-life.</td>
<td>I believe that physics knowledge helps us interpret seriously events in our daily life.</td>
</tr>
<tr>
<td>physics and occupational choice</td>
<td>Extent to which student use the information learned in the physics classroom for the futuristic work.</td>
<td>My career could be physicist/physics teacher/ engineer.</td>
</tr>
</tbody>
</table>

4.4 Data Collection and Analysis

The participants were asked to complete the perception questionnaire, to measure their pre-perceptions towards the computer-simulated physics experiment on liquid pressure, and the physics attitude questionnaire, to measure attitude towards physics lessons, for 15 minutes. After completing the instrument, they were exposed to interact dependently with the experiment for 25 minutes. After completing the experiment, the students' post-perceptions were examined by the same questionnaire for 10 minutes. The statistical data techniques selected for analyzing students' science motivation was repeated-measures MANOVA in SPSS to compare effect of intervention considering gender (female/male) and time (pre-test/post-test). In an addition, Pearson’s correlation was used to investigate relationship between physics attitude (interest in physics lesson, understanding and learning physics, importance of physics in real-life, physics and occupational choice) and their perceptions (flow, enjoyment, perceived learning, perceive ease of use, perceive of usefulness, perceive of satisfaction).

5. Results

The MANOVA indicated no significant main effect for gender (Wilks' lambda=0.875, F (6, 33) =0.783, p=0.589, partial \( \eta^2 = 0.125 \)). There was no significant difference on perceived towards learning science through computer simulation between females and males. The univariate results on gender revealed none of the six subscales on perceived towards learning science through computer simulation reached a statistical significance between females and males. That is both females and males performed indifferently with regard to perceive learning (PL), flow (Fl), enjoyment (Ej), perceive ease of use (PE), perceive of usefulness (PU), perceive of satisfaction (PS). Also, there was no significant interaction effect between gender and time (Wilks' lambda=0.915, F (6, 33) =0.795, p=0.085, partial \( \eta^2 = 0.090 \)). This means that the learning module has similar effects on perceived towards learning science through computer simulation for females and males. However, there was a significant main effect for time (Wilks' lambda=0.475, F (5, 40) =5.201, p<0.005, partial \( \eta^2 = 0.525 \)). The multivariate eta squared, \( \eta^2 \),
indicates the effect size, and a value of 0.525 means that about 52.5% of multivariate variance of the dependable variables was associated with time. The results of the univariate test for females and males students are summarized in Table 3.

Table 3: The students' perceptions towards computer-simulated physics experiment by time and univariate MANOVA.

<table>
<thead>
<tr>
<th>Dimension</th>
<th>Time</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Pre-test (3.16)</td>
<td>Post-test (2.09)</td>
<td>F</td>
<td>Sig.</td>
<td>η²</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Perceive learning (PL)</td>
<td>10.82</td>
<td>13.85</td>
<td>21.170</td>
<td>0.000***</td>
<td>0.385</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Flow (Fl)</td>
<td>12.86 (3.50)</td>
<td>15.62 (2.549)</td>
<td>12.322</td>
<td>0.001**</td>
<td>0.245</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Enjoyment (Ej)</td>
<td>8.54 (2.47)</td>
<td>9.82 (2.11)</td>
<td>4.647</td>
<td>0.038*</td>
<td>0.109</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Perceive ease of use (PE)</td>
<td>8.06 (2.23)</td>
<td>9.56 (1.93)</td>
<td>10.263</td>
<td>0.030*</td>
<td>0.213</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Perceive of usefulness (PU)</td>
<td>8.69 (2.40)</td>
<td>9.94 (1.71)</td>
<td>6.230</td>
<td>0.017*</td>
<td>0.141</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Perceive of satisfaction (PS)</td>
<td>8.74 (2.73)</td>
<td>10.79 (1.96)</td>
<td>14.158</td>
<td>0.000***</td>
<td>0.227</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. *p < 0.05; **p < 0.01; ***p < 0.001

As can be seen in Table 3, The univariate MANOVA on the six dimension scores of perceived towards learning science through computer simulation were significant differences across time, from pre-test to post-test. The univariate results revealed a significant effect on PL (F₁,₃₈ = 21.170, p < 0.05, partial η² = 0.385), Fl (F₁,₃₈ = 12.333, p < 0.05, partial η² < 0.245), Ej(F₁,₃₈ = 4.647, p < 0.05, partial η² = 0.109), PE (F₁,₃₈ = 10.263, p < 0.05, partial η² = 0.213), PU (F₁,₃₈ = 9.940, p < 0.05, partial η² = 0.141) and PS (F₁,₃₈ = 14.158, p < 0.05, partial η² = 0.227). According to aforementioned results, the overall result suggested that the increase of perceived towards learning science through computer simulation regarding perceive learning, flow, enjoyment, perceive ease of use, perceive of usefulness and perceive of satisfaction from the pre-test to post-test was homogeneous both females and males after participating with the computer simulation. That is, there was no effect of gender difference on perceived towards learning physics, through computer simulation learning.

For examining correlation between their physics attitudes and perceptions towards the experiment, Table 4 shows Pearson’s correlation of interest in physics lesson (IPL), understanding and learning physics (ULP), importance of chemistry in real-life (IPR), physics and occupational choice (POC), and perceive learning (PL), flow (Fl), enjoyment (Ej), perceive ease of use (PE), perceive of usefulness (PU), perceive of satisfaction (PS). Mean and standard deviation are also presented in Table 3.

Regarding Pearson’s correlation analysis of each variable, the result showed that there were no significant correlation between students' physics attitudes and perceptions towards the computer-simulated physics experiment. Thus, the result implied that the experiment could be used for all students even if they have a negative or positive attitude toward physics lessons.
Table 4: Descriptive and correlation for Attitude toward physics lesson and perceptions.

<table>
<thead>
<tr>
<th>Variable</th>
<th>IPL</th>
<th>ULP</th>
<th>IPR</th>
<th>POC</th>
<th>PL</th>
<th>Fl</th>
<th>Ej</th>
<th>PE</th>
<th>PU</th>
<th>PS</th>
</tr>
</thead>
<tbody>
<tr>
<td>IPL</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ULP</td>
<td>0.56**</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IPR</td>
<td>0.02</td>
<td>0.26</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>POC</td>
<td>0.25</td>
<td>0.30</td>
<td>0.09</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PL</td>
<td>-0.09</td>
<td>-0.22</td>
<td>0.07</td>
<td>-0.27</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FL</td>
<td>0.10</td>
<td>-0.07</td>
<td>0.09</td>
<td>-0.06</td>
<td>0.55**</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EJ</td>
<td>-0.03</td>
<td>-0.12</td>
<td>-0.00</td>
<td>-0.16</td>
<td>0.64**</td>
<td>0.79**</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PE</td>
<td>0.01</td>
<td>-0.23</td>
<td>-0.10</td>
<td>-0.10</td>
<td>0.26</td>
<td>0.53**</td>
<td>0.33*</td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PU</td>
<td>0.05</td>
<td>-0.16</td>
<td>0.23</td>
<td>-0.25</td>
<td>0.48**</td>
<td>0.70**</td>
<td>0.65**</td>
<td>0.45**</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>PS</td>
<td>-0.12</td>
<td>-0.10</td>
<td>0.14</td>
<td>-0.06</td>
<td>0.53**</td>
<td>0.52**</td>
<td>0.59**</td>
<td>0.30</td>
<td>0.70**</td>
<td>-</td>
</tr>
<tr>
<td>Mean</td>
<td>15.38</td>
<td>28.03</td>
<td>15.65</td>
<td>10.43</td>
<td>13.58</td>
<td>15.48</td>
<td>9.80</td>
<td>9.58</td>
<td>9.93</td>
<td>10.73</td>
</tr>
<tr>
<td>SD</td>
<td>2.77</td>
<td>4.71</td>
<td>3.17</td>
<td>2.51</td>
<td>2.09</td>
<td>2.54</td>
<td>2.11</td>
<td>1.93</td>
<td>1.71</td>
<td>1.96</td>
</tr>
</tbody>
</table>

Note. *p < 0.05, **p < 0.01

6. Discussions

This study reports an impact of an innovative technological tool for physics instruction, a computer-simulated experimentation on liquid pressure, for promoting students’ perceptions and attitude towards physics lessons. This tool was developed by emphasizing the interplay between macro- and micro-scale representations to induce cognitive process on construction of conceptual understanding and mental model. The result shows an increasing of students’ perceptions scores from pre-test to post-test reached a statistically significant effect across the time, and gender different had no significantly effect on the increasing of perceptions. This finding could be argued that in physics concept made progress throughout their experiencing with the computer-simulated experimentation on liquid pressure. This indicates that the experimentation successfully helped students getting better perceptions for physics learning of liquid fluid. The result is consistent with the research findings that students performed better achievements with learning from computer simulation (Tuan Soha, et. al., 2010). Based on attitude theory, Zimbardo and Leippe (1991) stated that attitudes can be either negative or positive and these attitudes can affect both perceptions and behavior. However, this study indicated contradict with the theory that there was no effect of students' prior physics attitudes on their perceptions towards computer-simulated experimentation.

7. Conclusion

A macro-micro representation-based computer simulation for physics learning in liquid pressure has been developed to promote students' conceptual comprehension and motivate them to learn physics more meaningfully. To preliminary evaluate its effectiveness for physics teaching and learning, this study employed the physics computer simulation to students and findings revealed that (i) gender difference has no effect on students' perceptions towards learning of physics through macro-micro representation-based computer simulation, and (ii) their attitudes towards physics lessons have no effect on the use of the computer simulation for physics learning in liquid pressure. As such, it is clear that both females and males increased their perceptions on perceive learning, flow, enjoyment, perceive ease of use, perceive of usefulness, and perceive of satisfaction after interacting with the simulation. Moreover, their attitudes towards physics lessons cannot intervene perceptions when they learn from the computer simulation. These findings could be used as a basis to develop an alternative computer
simulation for promoting physics instruction by emphasizing an interplay of macro- and micro-scale representations.

Acknowledgements

This work was financially supported by the Institute for the Promotion of Teaching Science and Technology (IPST) and Graduate School, Khon Kaen University. The author would like to express gratefully acknowledge to Faculty of Education, Khon Kean University, for supporting this contribution.

References


Promoting Students’ Physics Motivation by Blended Combination of Physical and Virtual Laboratory Environment: A Result on Different Levels of Inquiry

Chakkrapan PIRAKSA\(^a\), Niwat SRISAWASDI\(^b\)*

\(^a\)Science Education Program, Faculty of Education, Khon Kaen University, Thailand
\(^b\)Division of Science, Mathematics, and Technology Education, Faculty of Education, Khon Kaen University, Thailand

*niwsri@kku.ac.th

Abstract: In science and technology education community, technology-based pedagogy in science learning has been mentioned its effectiveness for facilitating scientific inquiry in school science. As such, this study investigated an effect of inquiry-based learning process into a blended combination of hands-on microcomputer-based laboratory and computer-simulated laboratory on secondary school students’ physics motivation. Study participants were 66 eleventh-grade students of diverse learning abilities in a public school in Northeastern region of Thailand. They were measured intrinsic motivation (IM), career motivation (CM), self-determination (SDT), self-efficacy (SEC), and grade motivation (GM) in physics learning by using a 25-item questionnaire both before and after participating the intervention. To evaluate the intervention, repeated-measures MANOVA was performed to examine its effects regarding type of inquiry (open- and guided inquiry) and time (pre- and post-test). The results showed that students’ physics motivation for pre- and post-test were significantly different and their motivation were improved after participating with blended combination of hands-on microcomputer-based laboratory and computer-simulated laboratory for both types of inquiry. This evidence indicated that inquiry-based physics learning with the blended laboratory environment (physical and virtual lab) influenced students’ progression of physics motivation. As such, blended combination of hands-on microcomputer-based laboratory and computer-simulated laboratory could be considered as a pedagogic technology-based laboratory environment for teaching and learning of science by inquiry.

Keywords: Sound wave interference, Microcomputer Based Laboratory (MBL), Computer simulation, and Motivation

1. Introduction

Several students have experienced difficulty in physics course due to misconceptions in many physics contents (Singh, Singh, Kumari, and Kumar, 2011), especially properties of sound wave. By the nature, properties of sound wave involve reflection, interference, reflection, diffraction, and also propagation of sound wave. Meanwhile, sound wave interference is invisible, complicated, and boring (Hola, 2007). Normally, teachers’ teaching in a regular classroom can encourage students to succeed in school and unable to motivation to learn as interact in complex ways to lead learning (Schunk, 2005). Teacher’s teaching approaches also depress motivation of students and decrease students’ learning performance in physics. Recently, most researchers have been concentrated on the scientific conceptions. The issues of motivation to learn physics has been attended in respect of science achievement and scientific conceptions (Hamzah and Mdzain, 2010). However, it’s difficult to achieve this ultimate goal because many learners are treated with a lack of motivation to learn science (Glynn, Brickman, Armstrong, and Taasoobshirazi, 2011).
In the recent year, students are educating a shift from passive sitting and listening to a more dynamic learning experience. Several active-teaching methods are introduced to solve those problems. Inquiry-based approach is a verity of instructional methods to apply with high school students. Science educators also confirmed methods of inquiry are more effective and valued to both teaching and learning (Guzey and Roehrig, 2009; Sadeh and Zion, 2011). The use of inquiry-based approaches is strongly subscribed to teaching and learning of science (Minstrell and VanZee, 2000), student-centered, providing students with opportunities to formulate and conduct their own scientific investigations (Singer, Marx, and Krajcik, 2000). Scientific inquiry tasks play an important role for students in the process of conceiving scientific problems and questions, formulating hypotheses, designing experiments, gathering and analyzing data, and drawing conclusions (Hofstein, Navon, Kipnis, and Naaman, 2005). The researchers also revealed that the cookbook-laboratory activities do not promote the development of students’ higher order thinking skills. On the other hand, in inquiry-based laboratory students are more associated with, and usually have positive attitudes regarding their laboratory experience (Abd-El-Khalick and Akerson, 2004).

On top of that, this learning process is a wide range of efficient technological environments and applications including animations, simulations and modeling tools, microcomputer-based laboratories (MBL), intelligent tutoring systems, web resources and environments, spreadsheets, scientific databases, for instance, in the science education community of practice. Using as tools, MBL and computer simulation are subject to introduce students’ cognitive development and result in students’ positive response (Hola, 2007) because they facilitated to understand the scientific conceptions that confront them (Mulder, Lazonder, and Jong, 2011; Russell, Lucas, and McRobbie, 2003). It was not until third decade ago, MBL was reported to understand and integrate learners the sophisticated topics of physics including temperature probe, heat energy (Russell et al., 2003), and properties of sound wave (Gunhaart and Srisawasdi, 2012). Furthermore, the capacity of MBL enable learners to immediately transform data from each experiment into graph, the most powerful form of presentation. Learners will be engaged a construct and had conceptualized change after all. In the interval, computer simulations are examined to be the most technically complicated option for offering various benefits for the teaching and learning of science (Blake and Scanlon, 2007). For this reason, a well-designed computer simulation used within MBL as educational technology and inquiry learning as instructional process can be very effective in promoting meaningful learning in scientific concepts (Bell and Trundle, 2008).

This research utilized both tools to engage learners a meaningful learning of sound wave interference. Conceptualized change was expected to achieve by measuring five components involving Intrinsic motivation (IM), Career motivation (CM), Self-determination (SDT), Self-efficacy (SEC), and Grade motivation (GM). Inquiry types were examined as dependent variables for motivation.

2. Literature Review

2.1 Inquiry

In Thailand, instructors popularly recommend to use inquiry-based learning as one of many instructional strategies to implement in science education. Theoretically, inquiry-based learning is defined as the creation of a classroom where students are engaged in essentially open-ended, student-centered, hands-on activities involving asking questions about the world around them, gathering evidence, and providing explanations (Colburn, 2000). It is restricted that solely activities, e.g. building a model of an atom, cannot be referred to inquiry-based learning if they are conducted in the absence of research questions as a part of inquiry process. The inquiry-learning literature tends to be more closely associated with the acquisition of science process skills or the scientific thinking and reasoning patterns that scientists use to construct (Bunterm et al., 2014).

Researchers typically discriminate between different levels of inquiry-based learning depending on the amount of specific instructions given to students. (Buck, Bretz, and Towns, 2008). Buck (2008) proposed a fifth-level model. At the first level (Confirmation), the problem, procedure, analysis, and correct interpretations of the data are all provided to the students. At the second level (Structured inquiry), the laboratory manual provides the problem, procedures, and analysis by which students can discover relationships or reach conclusions that are not already known from the manual. At
the third level (Guided inquiry), the laboratory manual provides the problem and procedures, but the methods of analysis, communication, and conclusions are for the student to design. At the fourth level (Open inquiry), the problem and background are provided, but the procedures/design/methodology are for the student to design, as are the analysis and conclusions. At the highest level (Authentic inquiry), the problem, procedures/design, analysis, communication, and conclusions are for the student to design.

This investigation compared two kinds of inquiry-based processes: guided versus open inquiry. Learners will be engaged to have a construct with providing the problem and procedures for guided inquiry but providing just the problem for open inquiry. Sadeh and Zion (2011) examined the influence of these two different inquiry learning processes on the attitudes of Israeli high school Biology learners toward their inquiry project. It is found that there were significant differences between the two groups. Learners were more satisfied and felt they gained benefits from implementing the project to a greater extent for open inquiry. On the other hand, they conducted more documentation for Guided inquiry.

Bunterm et al. (2014) examined the effects of guided against structured inquiry on secondary students’ learning of science with three schools in north-eastern Thailand. In comparison, students in the guided-inquiry condition showed greater improvement in both science content knowledge and science process skills. Any moment now, researchers have been subject inquiry-based learning using MBL and computer simulation as tools for conceptual understanding and change in physical science to middle and high school students, pre-service teachers to enhance learners’ meaningful learning in the area of scientific concepts (Gunhaart and Srisawasdi, 2012).

2.2 Microcomputer based Laboratory (MBL) and Computer simulation for science instruction

At this moment in time, computational technologies are increasing attention among science educators because of their potentials to support new variety of science classroom (Srisawasdi, 2008). MBL and computer simulation are taking participant their own prominent rules in thinking and reflecting learning input for an instructor and a conceptual construct respectively. Serving as alternative software for teaching assistant tool, MBL is widely used for instructional activities to stimulate students’ curiosity as a learning motivator, develop students’ scientific skills, foster collaborative network, understand in scientific concepts, and establish students’ cognitive construct (Srisawasdi & Kroothkaew, 2014).

Additionally, Redish, Saul, and Steinberg (1997) investigated that active-engagement tutorials using MBL equipment were replaced for traditional problem-solving recitations in introductory calculus-based mechanics classes for engineering students at the University of Maryland. Two specific tutorials, on the concept of instantaneous velocity and Newton’s third law, were performed with eleven lecture classes taught by six different teachers with and without tutorials. Classroom achievement tests were probed by using standard multiple choice questions and a free-response final exam question. The result shows that the MBL tutorials originated in a remarkable improvement compared to the traditional recitations. Russell et al. (2003) designed and provided experiments with grade 11 physics classes of 29 students. The research distinguished the learners and illustrated the patterns of interactions in the MBL. Analysis of students’ discourse and actions identified kinematics in multiple ways. The finding is that MBL activities likely catalyzed students’ construction of understanding. Students were able to design the research questions, predictions, designing experiments, collecting data, and drawing conclusions.

In addition, Gunhaart and Srisawasdi (2012) used MBL as a tool for scientific thinking and computer simulation as a cognitive tool for conceptual learning to improve the construction of physics conceptual understanding on properties of sound wave at macroscopic (observable) level. The results show achievement caused them importantly obtaining a better conceptual score at the end of their learning. In addition, the qualitative analysis suggests the students had changed their conceptual understanding on physics of sound wave properties in three characteristics including differentiation, class extension, and reconceptualization. Srisawasdi (2012) introduced MBL and computer simulation to cover basic science concepts including three physical science activities: air resistance of falling objects, heat of fusion for ice and photosynthesis and respiration for 26 second year pre-service teachers in Thailand. Results indicate that all the groups did not perceive differently the goal and the support of computerized science laboratory. The highest attitude group realized the ease of use, self-learning, and value greater than the medium and the low attitude groups, but the medium attitude group possessed the most satisfaction with the laboratory.
2.3 Motivation

Motivation stands for an internal state that activates, guides, and maintains behavior, students’ drive to learn and achieve to their potential at school. There are five components of motivation to learn science (Glynn et al., 2011) including: First, intrinsic motivation (IM) involves learning science for its own sake. Second, career motivation (CM) included the relevance of science to one’s career is a leading theme in students’ explanation. Third, self-determination (SDT) as dimension refers to the control students’ believe that they have over their learning of science. Forth, self-efficacy (SEC) is an achievement by predisposing students to work harder, persist longer, and overcome barriers when pursuing academic goals and finally, grade motivation (GM) as the students’ competition often associate with grade.

In term of science education, motivation to learn is the engagement related to the achieving goals, students’ understanding of science and the activation of strategies for action (Lee & Brophy, 1996). Due to the relationship between motivation, cognitive engagement and conceptual change, motivation to learn is a particular issue to concern in science education. Pintrich, Marx, and Boyle (1993) have suggested that the construction of new knowledge in science is strongly influenced by prior knowledge, conceptions gained prior to formal learning. Consequentially, conceptual change is much tougher because it requires new information to engage at an adequately deep level to recognize conflicts between existing understanding and new information (DeBacker and Nelson, 2000). Confirmation persuades that decisions to engage in endeavored learning might be affected by individual students’ motivation including engaging goals in an activity, beliefs in abilities and the nature of the task, and valuing of the task (Miller, Greene, and Montalvo, 1996; Nolen and Haladyna, 1990). In addition, learning environment, especially laboratory maybe the key factors affecting motivation differences.

3. Methodology

3.1 Research Objective

this study aims to investigate students’ physics motivation delivered in blended combination of hands-on microcomputer-based laboratory and computer-simulated laboratory for physics learning of sound wave. Specifically, the main research questions for this study was that do the students engaged in blended combination of hands-on microcomputer-based laboratory and computer-simulated laboratory perform significantly better by students’ physics motivation?

3.2 Study Participants

The total of 66 students-respondents in their eleventh grade (16-17 years old) were recruited in this present study. They were divided into two experimental groups which received different learning process in blended combination of hands-on microcomputer-based laboratory and computer-simulated laboratory: open-inquiry laboratory learning (N=31) and guided-inquiry laboratory learning (N=35) groups. Both groups were assigned to learn a physics lesson on sound wave. The researchers conducted an informal interview with physics teacher in two regular classes, and the results showed that all of students have basic skills on using computer. However, all of them have never experience yet using hands-on microcomputer-based laboratory and computer-simulated laboratory in physics class.

3.3 Instrument

In this study, a 25-item science motivation questionnaire was used to measure students’ motivation to learn physics on five subscales: intrinsic motivation (IM), self-determination (SDT), self-efficacy (SEC), career motivation (CM), and grade motivation (GM) (Glynn et al., 2011). The questionnaire was
originally developed by Glynn et al. (2011) and then adapted into Thai version to assess students’ motivation to learn science. From 25 items English version, the translation an identical version in Thai was constructed and Cronbach’s alpha of Thai version were 0.79, 0.81, 0.89, 0.81 and 0.85 for IM, SDT, SEC, CM and GM respectively (Srisawasdi, submitted).

3.4 Data Collection

For investigating students' physics motivation in whether they perform inquiry-based learning process with a blended combination of hands-on microcomputer-based laboratory and computer-simulated laboratory on sound wave phenomena. The study participants were asked to response the 25-item 5-point Likert-scale questionnaire for 10 minutes at both before and after interacting with the blended lab. On each item, respondents were assigned to rate how much the respondent agree with into five scale, from 1-strongly disagree to 5-strongly agree. In the blended lab class, both groups participated physics learning of sound wave through inquiry-based learning process for 480 minutes. Figure 1 and 2 illustrate example of blended lab activity for physics learning of sound wave.

![Figure 1. Illustrative example of classroom learning activity through hands-on MBL guided (Left) and open (Right) inquiry laboratory](image)

![Fig. 2 Illustrative interface the bending interference of sound wave simulation (obtained from PhET) for computer-simulated guided and open inquiry laboratory](image)

The statistical data techniques selected for analyzing students' perceptions was repeated-measures MANOVA in SPSS 21.0.

4. Results

The results for the repeated-measures MANOVA indicated significant main effect for different levels of inquiry (guide- and open inquiry) (Wilks’ lambda=0.755, F (5, 60) = 3.887, p = 0.004, partial η² = 0.245). There was significant difference on students' physics motivation between guided- and open-inquiry learning process. According the significance, the univariate results was performed and it
revealed that all of the five subscales on physics motivation reached a statistically significant difference between guided- and open-inquiry learning process. That is, both guided- and open inquiry in blended combination of hands-on microcomputer-based laboratory and computer-simulated laboratory performed differently with regard to IM, CM, SDT, SEC, and GM. In additions, there was a significant interaction effect between different levels of inquiry and different times measured (pre- and post-test) (Wilks' lambda = 0.717, \( F(5, 60) = 4.738, p = 0.001 \), partial \( \eta^2 = 0.283 \)). This means that different levels of inquiry had similar effects on students' physics motivation in the blended lab. In addition, there was a significant main effect for different time measured (Wilks' lambda = 0.483, \( F(5, 60) = 12.855, p < 0.000 \), partial \( \eta^2 = 0.517 \)). This suggests that, on average, the students' physics motivation have changed over inquiry-based learning experience with blended combination of hands-on microcomputer-based laboratory and computer-simulated laboratory. Univariate analyses of variances on each subscale were conducted as follow-up tests to the one-way MANOVA. The results of the univariate test regarding different time measured are summarized in Table 1.

<table>
<thead>
<tr>
<th>Subscale</th>
<th>Time</th>
<th>F</th>
<th>Sig.</th>
<th>( \eta^2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intrinsic motivation (IM)</td>
<td>Pre-test</td>
<td>17.58 (3.123)</td>
<td>19.85 (2.562)</td>
<td>29.920</td>
</tr>
<tr>
<td>Career motivation (CM)</td>
<td>Post-test</td>
<td>17.17 (3.580)</td>
<td>19.92 (2.668)</td>
<td>39.803</td>
</tr>
<tr>
<td>Self-determination (SDT)</td>
<td>Pre-test</td>
<td>17.23 (2.971)</td>
<td>18.21 (2.551)</td>
<td>5.675</td>
</tr>
<tr>
<td>Self-efficacy (SEC)</td>
<td>Post-test</td>
<td>14.55 (3.398)</td>
<td>15.91 (3.703)</td>
<td>11.467</td>
</tr>
<tr>
<td>Grade motivation (GM)</td>
<td>Pre-test</td>
<td>19.33 (2.879)</td>
<td>21.21 (2.551)</td>
<td>24.289</td>
</tr>
</tbody>
</table>

As displayed Table 1., The univariate MANOVA on the five subscale scores of physics motivation were significant differences across time, from pre-test to post-test. The univariate results revealed a significant effect on IM (\( F_{1,64} = 29.920, p < 0.001 \), partial \( \eta^2 = 0.319 \)), CM (\( F_{1,64} = 39.803, p < 0.001 \), partial \( \eta^2 = 0.383 \)), SDT (\( F_{1,64} = 5.675, p < 0.05 \), partial \( \eta^2 = 0.081 \)), SEC (\( F_{1,64} = 11.467, p < 0.01 \), partial \( \eta^2 = 0.152 \)), and GM (\( F_{1,64} = 24.289, p < 0.001 \), partial \( \eta^2 = 0.275 \)). According to aforementioned results, the overall result suggested that the increase of physics motivation regarding intrinsic motivation, career motivation, self-determination, self-efficacy, and grade motivation from the pre-test to post-test was homogeneous for both guided- and open-inquiry learning process after participating with blended combination of hands-on microcomputer-based laboratory and computer-simulated laboratory. That is, there was effect of different levels of inquiry on students' physics motivation for learning with blended combination of hands-on microcomputer-based laboratory and computer-simulated laboratory.

5. Discussion

This research reports an effect of innovative teaching and learning of physics, inquiry-based learning process in a blended combination of hands-on microcomputer-based laboratory and computer-simulated laboratory, on promoting students' physics motivation. The result indicated an increasing of students' physics motivation scores considering from before and after participating with the intervention. This finding could be discussed that inquiry types subjecting to five subscales of physics motivation are shown in Table 1. In cases of guided inquiry with innovatively effective tools of MBL and computer simulation persuaded students to particularly focus on given content, interference of sound wave and delivered them opportunities to construct knowledge with team groups (Zion, Cohen, and Amir, 2007) and prospectively achieved conceptual change after students attended this learning process. This indicated that learners were more satisfied to explore a knowledge construct themselves with MBL. This study showed time as the main effect of the learning process to the students’ motivation on sound wave interference. According to the results, there were statistical significant effect in all subscales of the students’ physics motivation. This result implied that the learning process which were MBL and computer simulation could motivate the student to learn physics. Due to learning using both MBL and computer simulation, the students had higher scores of the
motivation. Considering to MBL, this method allows students to learn through actual laboratory using technology as tools supporting their learning process (Russel, 2003) and computer simulation by which they had background of sound wave interference that invisible in micro level (Srisawasdi, 2008). The result consistent with the research findings that students perform better in physics concepts with learning from integrating of MBL and computer simulation (Gunhaart and Srisawasdi, 2012). A possible explanation for why learners made develop on physics motivation from before to after is that the teaching and learning could induce learners into the problem solving (Russell et al., 2003; Thornton and Sokoloff, 1989).

Considering different levels of inquiry (open- and guided inquiry), the findings introduced to acquire more effective process. One of the best findings was that open-inquiry laboratory learning was more effective learning process to motivate student in physics learning of sound wave than that of guided-inquiry laboratory learning. The results showed there was a significant difference for all of motivational subscales in both inquiry levels. This evidence is consistent with the claims that the inquiry are a well designed learning process for science learning and can engage mindful investigation in doing science (Bunterm et al., 2014; Sadeh a Zion, 2011). Moreover, open-inquiry learning through computer simulation affected students revising unscientific understanding and improving their physics outcomes (Srisawasdi, 2014). Also, Srisawasdi (2012) has mentioned that hands-on microcomputer-based laboratory support improvement of attitude and perception toward learning. Therefore, this implied that using inquiry-based could support the students’ leaning in affective domain such as motivation. Especially, open-inquiry laboratory learning process where student have opportunity to design, collect and analysis data, discuss with peers, make conclusion and communicate findings by their own way delivered them a novel learning process of science can motivate to learn physics greater than prescribed physics experimentation.

6. Conclusion

This study investigated an effect of inquiry-based learning process into a blended combination of hands-on microcomputer-based laboratory and computer-simulated laboratory on secondary school students’ physics motivation. After implementing the intervention, the results show that; (i) both guided- and open-inquiry learning process in blended combination of hands-on microcomputer-based laboratory and computer-simulated laboratory improved students’ intrinsic motivation, career motivation, self-determination, self-efficacy, and grade motivation towards physics learning across time; and (ii) open-inquiry laboratory learning process was more effectively to enhance students' physics motivation than the guided inquiry. To this end, blended combination of physical, hands-on microcomputer-based laboratory, and virtual, computer-simulated laboratory, environment could be used to motivate student in learning of physics by inquiry. However, to address students’ conceptual learning performance we are going to investigate how to use the inquiry-based learning process through blended lab for facilitate development of mental model and ability of scientific reasoning.

Acknowledgements

This work was financially supported by the National Research University (NRU) program of Khon Kaen University, and Graduate School, Khon Kaen University. The author would like to express gratefully acknowledge to Faculty of Education, Khon Kean University, for supporting this contribution.

References


An Experimental Study on the Effects of an Online Student-Constructed Tests Learning Activity

Fu-Yun YU* & Chia-Ling SU
Institute of Education, National Cheng Kung University, Taiwan
*fuyun.ncku@gmail.com

Abstract: While the learning potential of student-constructed tests for the promotion of knowledge integration and elaboration has been suggested, its learning effects warrant further empirical examination. Three fifth-grade classes (N=76) participated in this study for nine consecutive weeks. A one-group pre-post experimental research design was used, and an online student-constructed tests learning system was adopted to support elementary students’ science learning. The results from the paired t-tests found significant increase in students’ attitudes toward science and science learning motivation as a result of the incorporated activity. Yet, no significant differences were found in students’ use of cognitive and metacognitive strategies after the activity. Based on the collected data, suggestions for instructional implementations are provided.

Keywords: Experimental study, learning effects, online learning system, primary school settings

1. Introduction

The learning benefits of student-generated questions (hereafter name SQG) have been well established empirically. In general, empirical evidence accumulated since the 1960s provides a solid basis for its effects on enhancing understanding, academic achievement, motivation, question-generation abilities, the use of cognitive and metacognitive strategies, problem-solving abilities and attitudes toward the subject matter studied (Brown and Walter, 2005; Chin, Brown and Bruce, 2002; Dori and Herscovitz, 1999; English, 1997; Perez, 1985; Rosenshine, Meister and Chapman, 1996; Yu and Liu, 2008).

Recently, researchers have experimented the idea of engaging students in constructing a test and found promising evidence for its potential. Specifically, data on students’ perceptions found that students’ preference to and perceptions of student-constructed tests (SCT) and teacher-constructed tests (TCT) were statistically significant at $p < .01$ with a considerable proportion of students preferring SCT as the approach for assessing their learning and regarding SCT as a better approach for promoting learning (Yu, 2013). Descriptive data analyzed further highlighted the potential of SCT for the promotion of knowledge integration (Yu and Su, 2013a) and knowledge elaboration (Yu and Su, 2013b).

Constructing “tests” is different from constructing questions, and it would direct attention to additional criteria. Since a more holistic view of the study content may be obtained (Yu and Su, 2013a), and cognitive processes of different nature and intensity may be mobilized, the learning effects of SCT was the focus of this study. To provide comprehensive information about the observed phenomena in educational context, the learning effects on both cognitive and affective (specifically, the use of cognitive and metacognitive strategies, attitudes toward science and science learning motivation) are examined in this study.

2. Methods

2.1 Participants and instructional content
Students from three intact fifth-grade classes (N=76) taught by the same science teacher were briefed about the purpose of the introduced online learning activity (i.e., support of their science learning; the promotion of higher-order thinking skills, including self-monitoring comprehension level when attending lectures, grasping the main ideas of the study content; evaluating self- and peers’ learning by constructing a set of questions of appropriate scope and difficulty) and invited to participate. All students at the participating schools started taking computer classes when they were at the 3rd grade and so had basic computer skills needed to carry out the activity.

Two science units were covered during the study. The first unit is about “how heat affects matter” with three lessons that cover topics including the changes of matter after heating, heat transfer, and insulation. The second unit is on “air and burning” with three lessons. Topics covered include the characteristics of oxygen and carbon dioxide, their uses in daily life and their relationship to burning matter; three elements of combustion; and fire extinguishing and the fire prevention.

2.2 Implementation procedures

The implementation procedure is delineated in Figure 1. A pilot study involving one fifth-grade class in the participating school (N=28) was conducted to ensure that the planned procedures and time allocation for various activities were appropriate prior to the actual study. Data on participants’ cognitive and metacognitive strategy use, attitudes toward science and science learning motivation was collected prior to the commencement of this study.

![Figure 1. Experimental procedures of this study](image)

This study took place right after the school-wide first-term exam. For this study, three of the most frequently encountered question types in primary schools in Taiwan were chosen for the learning activity—true/false, matching and multiple-choice questions. For the duration of the study (i.e., nine consecutive weeks in total), as a routine, participating students would head to the computer lab after attending three 40-minute instructional sessions on science in their respective class. To equip participants with essential knowledge and skills associated with the engaged tasks, three sessions were reserved for training prior to the study. During the training session, quality criteria frequently associated with SQG and SCT and basic principles for item writing for each of the chosen question types were introduced and explained. In addition, the operational procedures for the adopted system were demonstrated, followed up by students’ hands-on spaced practice activities.

For each of the following six weeks, at the beginning, whole-class feedback on student performance at the previous SCT activity was arranged with reference to SCT criteria (e.g., covering...
all main topics, appropriateness of test difficulty level in general, appropriateness of coverage and representation of all main topics). Afterwards, students were directed to construct a test around the science content covered in the prior three instructional sessions by composing a minimum of five question items consisting of at least two out of the three chosen question types. A post-session questionnaire were disseminated to participants for individual completion after the study.

2.3 The Online SCT Learning System

An online learning system supporting associated activities of the study was adopted. Students in this study had access to test-construction and test-view functions of the adopted system. To construct a test, students first design the overall structure of a test in terms of the number and scoring scheme of each question type. Second, students generate questions out of any of the three question types of their choice. After satisfying a number of questions have been generated, students then view and select individual questions to be included in the test at work. Finally, students can determine and re-arrange the relative sequence of questions both within and among question types before submission.

To promote learning by permitting students to learn from observing peer’s work, an observational learning space—test-viewing was created and made accessible at the last 15 minutes of each online learning activity.

2.4 Measurement instrument

Three instruments were used in the study to test the learning effects of online student-constructed tests. First of all, Hung’s (2002) “Learning Strategy Use Scale” was adopted. The scale consisted of two parts: “Cognitive Strategies Use Scale” (18 items) and “Metacognitive Strategies Use Scale” (24 items). The former appraises students’ use of rehearsal, elaboration and organization learning strategies, and the latter reveals students’ activation of metacognitive strategies for cognition regulation, such as planning, monitoring, revising and evaluating one’s actions and reasoning while learning. All items were rated on a 6-point Likert scale, with corresponding verbal descriptions ranging from “no consistency” through “very inconsistent,” “somewhat inconsistent,” “somewhat consistent,” “very consistent,” to “complete consistency.” The internal consistency reliability calculated after this study was .92 and .94 for the “Cognitive Strategies Use Scale” and “Metacognitive Strategies Use Scale,” respectively.

Second, “Attitude toward Science in School Assessment” developed by Germann (1988) was adopted to measure students’ attitudes toward science. To ensure that the instrument was translated appropriately and adequately, back translation technique was used. Results from the exploratory factor analysis and Cronbach’s α with a group of fifth-grade students (N=30) by Tsai (2010) substantiated its validity and reliability. The instrument consisted of fourteen Likert-scale items. Each statement was rated on a five-part discrete scale, with corresponding verbal descriptions ranging from “strongly disagree” through “disagree,” “no-opinion,” “agree,” to “strongly agree.” The Cronbach’s alpha values calculated after the study (N=149) was .88.

Finally, Hung’s (2002) “Science Learning Motivation” was adopted for this study. The scale consisted of 14 items and used a six-point Likert scale (ranging from 1=no consistency to 6=complete consistency). The scales validated by a group of 303 sixth-graders evidenced good validity. The Cronbach's alpha values calculated after the study (N=149) was .94.

3. Results

As shown in Table 1, after exposed to the SCT activity, students not only activated more of cognitive and metacognitive strategy while learning science, but also formed better attitudes and exhibited heightened motivation toward science. Nevertheless, the results from one-group paired t-tests found significant differences only in attitudes toward science and science learning motivation, but not in the cognitive domains.
<table>
<thead>
<tr>
<th>Variables</th>
<th>n</th>
<th>M(SD)</th>
<th>t</th>
<th>η²</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cognitive strategy use</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pretest</td>
<td>76</td>
<td>3.86(0.95)</td>
<td>-1.12</td>
<td>.02</td>
</tr>
<tr>
<td>Posttest</td>
<td></td>
<td>4.01(1.20)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Metacognitive strategy use</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pretest</td>
<td>76</td>
<td>4.00(0.93)</td>
<td>-1.56</td>
<td>.03</td>
</tr>
<tr>
<td>Posttest</td>
<td></td>
<td>4.17(1.19)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Attitude toward science</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pretest</td>
<td>76</td>
<td>4.41(1.27)</td>
<td>2.89*</td>
<td>.10</td>
</tr>
<tr>
<td>Posttest</td>
<td></td>
<td>4.76(1.09)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Science learning motivation</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pretest</td>
<td>75</td>
<td>4.10(1.14)</td>
<td>3.13*</td>
<td>.17</td>
</tr>
<tr>
<td>Posttest</td>
<td></td>
<td>4.47(1.15)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*p < .05

4. Discussion and Conclusion

Preliminary studies on the potential of SCT supported its effects on knowledge integration and elaboration (Yu and Su, 2013a; Yu and Su, 2013b). The current study extended prior studies by substantiating its affective effects. By allowing students to construct questions around the study materials they regard as important and relevant, to allocate different weighting among different study topics and to decide the relative sequence of question items within and among question types, SCT in essence is more in alignment with what constructivism, self-regulation and self-determination theories accentuate. As a result, as found in this study, exposing students to SCT helped to increase students’ attitudes toward science and science learning motivation.

However, the current study failed to find SCT helped to promote the use of cognitive or metacognitive strategies. Through in-depth analysis of the current study and prior studies, some possible explanation for the unconfirmed results are rendered. First of all, the current study involved fifth-grade students (average age=11), who just reached Piaget’s formal operational cognitive development whereas prior studies involved university students, who should be mentally more prepared and ready for the whole range of tasks involved in SCT. Second, participants in this study were directed to construct tests around the study materials on a weekly basis, which may not be in its entirety. Unlike prior studies, SCT activity was arranged around the end of the semester where opportunities for interconnecting and integrating of all topics are provided.

Based on the findings of this study, it is suggested that instructors can engage students in SCT activity for the promotion of students’ affective development, in specific, attitudes and learning motivation toward the learned subject.

Acknowledgements

This paper was funded by research grant from the National Science Council, Taiwan (Project: Online student-generated tests learning system: Development, applicability and learning effects, NSC 102-2511-S-006-003-MY3).

References


Exploring the Effects of Student Question-Generation Strategy

Chun-Ping WU*a, Shu-Ling WUb & Ching-Chiu YEHc

*aDepartment of Educational Technology, TamKang University, Taiwan
bGraduate Institute of Human resource and eLearning Technology, National Hsinchu University of Education, Taiwan
cDepartment of Educational Technology, TamKang University, Taiwan
*cpwu303@gmail.com

Abstract: The purpose of this study is to explore the effect of student question generation strategy on students’ reported use of cognitive strategies and metacognitive strategies. Furthermore, the relationship between the SQG performance and academic achievement was also examined. A single group experiment was implemented for 7 weeks. Seventy-two junior high school students from two intact history classes were recruited. Participants were engaged in the SQG task followed by the peer-assessment activity. The finding supported the positive effect of the SQG on enhancing students’ use of cognitive strategies and metacognitive strategies. Additionally, students’ SQG performance was significantly correlated with their academic achievement, which was supported.

Keywords: Student Question Generation, cognitive strategies, metacognitive strategies, academic achievement

1. Introduction and Literature Review

Students’ ability to raise a good question relies on their use of existing knowledge to observe and interpret the newly learned content or phenomena. Therefore, it brings the needs to explore how to facilitate students in bridging the new content with their knowledge bases. The student question-generation (SQG) strategy, which is grounded on the constructivism and information processing theory, has gained more attention from the researchers and educators (i.e. Abramovich & Cho, 2006; Berry & Chew, 2008; Brown & Walter, 2005; van Blerkom, van Blerkom, & Bertscho, 2006; Yu & Wu, 2013). The question-composing and revising task could engage students in recalling, organizing or elaborating the newly learned content.

Specifically, the SQG process requires students to recall the content they just learned and identified important concepts and the concepts which their peers might be confused about. Those identified concepts could be used as the focus of the question. For example, while designing the multiple-choice question, students have to examine the interconnection among concepts and translate their understanding into the question stem in their own words or using appropriate examples. Additionally, while designing the correct answer and the three distractors, the question authors experience a micro problem-solving process (Yu, Liu, & Chan, 2005). They have to propose different solutions to the questions and examine and compare the solutions to ensure one best correct answer. Therefore, the question-generation process engages students in organizing, analyzing the learned contents, examining their understanding and misconception (Lee & Hutchison, 1998), and elaborating the contents in a meaningful way which helps to construct their schema (Bangert-Drowns, Hurley &Wilkinson, 2004; Herbert & Burt, 2004).

The SQG effects on enhancing students learning motivation, confidence, understanding of the learning materials, metacognition and so on have been supported in empirical studies. (Abramovich & Cho, 2006; Barlow & Cates, 2006; Belanich, Wisher,& Orvis, 2004; Berry & Chew, 2008; Brown & Walter, 2005; Choi, Land, & Turgeon, 2005;Dori & Herscovitz, 1999,2005; Fellenz, 2004; Ikuenobe, 2001; van Blerkom, van Blerkom, & Bertscho, 2006; Whiten, 2004; Wilson, 2004; Yu, 2005, 2009; Yu & Liu, 2005)
The purpose of this study is to further validate whether the above-mentioned SQG process would enhance students use of cognitive strategies and strategies in the context of junior high school’ history course. Furthermore, as suggested, the SQG might help students’ deep understanding of the learned content. The second purpose of this study is to examine whether students with better question-generation performance also performed better in the academic achievement tests.

2. Research Method

2.1 Research Design

Seventy-two junior high school students from two intact history classes taught by the same instructor were invited to participate in this study. A single group experimental design was implemented for the seven weeks. At the beginning, the purposes of the question-composing activity were explained to the participants followed by the training.

As suggested by theories and literature, the question composing task is difficult especially for those students without question composing experience (Yu, 2009), thus training on question posing is essential. Students who do not have knowledge of the quality criteria of a good question or are not familiar with the reasoning process of composing a question, might devote efforts to composing questions measuring the facts rather than higher level questions. Furthermore, without developing the schema of question-composing process, they might encounter difficulty in either translating the concepts into question stem or offering the groups of options that are highly related to the question stem. On the basis of the needs for the training, this study incorporated several components into training: the quality criteria of a good multiple-choice question stem and four options, which include one answer and three distractors, the reasoning process of question posing and revision, the explanation of the value of the SQG and hands-on practice followed by feedback.

During the intervention, the participants were required to compose two to three multiple-choice question items in accordance to each of the five instructional topics. A peer-assessment activity was conducted at the instructional session followed by the SQG activity. A whole-class feedback on SQG performance was provided and the peer’s comments collected during the peer-assessment activity were sent to the question-author.

2.2 Variables and Instruments

The examined cognitive strategies were defined as students’ reported use of rehearsal, elaboration and organization strategies while the metacognitive strategies were defined as their reported use of planning, monitoring and self-judgment during the learning process. These two variables were measured by the translated version of Motivated Strategies for Learning Questionnaire (MSLQ) (Garcia & Pintrich, 1995). The Cronbach’s α for cognitive (10 items) and metacognitive strategies (11 items) were 0.90 and 0.89 respectively. Students rated themselves on a seven point Likert scale from “not at all true of me” to “very true of me”. Scales were constructed by taking the sum of the scores of items that make up that composite construct of the scale.

Additionally, to examine SQG performance, all the questions were evaluated by two independent raters. The evaluation criteria were adopted from the index, proposed by Yu & Wu (2013) and were revised in accordance with the course instructor’s suggestions. The criteria included four dimensions: Importance, fluency, elaboration and cognitive Level.

To establish the inter-rater reliabilities, one third of students composed questions were randomly selected from 822 questions and evaluated by another independent rater (N = 274). The results of the inter-rater reliability were $r = 0.87, p < 0.01$, which proved to be satisfactory.

To examine students’ academic achievement in the five instructional units, students’ performance in the school tests were collected.
3. Results and Conclusions

3.1 Findings of the SQG Effects on Students’ perceptions

The descriptive statistics of the variables are listed in Table 1. It can be seen that the post-test scores of students’ reported use of cognitive strategies and metacognitive strategies (Mean=4.94, 4.74, respectively) are higher than the pre-test scores (Mean=4.52, 4.33, respectively).

Table 1: Descriptive statistics (N=72)

<table>
<thead>
<tr>
<th>Variable</th>
<th>Use of Cognitive Strategy</th>
<th>Use of Metacognitive Strategy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pretest Mean (SD)</td>
<td>4.52 (1.09)</td>
<td>4.33 (0.84)</td>
</tr>
<tr>
<td>Posttest Mean (SD)</td>
<td>4.94 (1.11)</td>
<td>4.74 (0.99)</td>
</tr>
</tbody>
</table>

The paired t-test result showed that the participants’ post-test score of students’ reported use of cognitive strategies is significantly higher than the pretest score. \( t=2.91, p = .005 \). Similarly, the participants’ post-test score of reported use of metacognitive strategies is significantly higher than the pretest score. \( t=4.12, p < .01 \). In other words, the participants’ reported use of cognitive strategies and metacognitive strategies were significantly enhanced after being engaged in the question-generation activity.

3.2 Findings of the relationship between SQG Performance and Academic Achievement

During the seven-week intervention, 822 questions were generated by 72 participants. As specified in the data analysis section, the questions were evaluated by the two raters using the pre-defined criteria and the ratings were adopted as the indicators for students’ question-generation performance.

The mean scores of students’ overall achievement, achievement in each unit and question-generation performance as well as the correlations among variables were presented in Table 2. As shown, the participants’ overall question-generation performance is significantly correlated with their achievement. Furthermore, in order to explore the relationship between students’ question-generation performance and their gained knowledge on each unit, the correlation analyses were conducted. The results show that participants’ question-generation performance in unit 1 and 2 is significantly correlated with their achievement scores gained in the test of unit1 and 2. Similar result was found in the unit 3 and 4. The hypotheses that students who generated questions of better quality tended to perform better in the achievement tests were supported in this study. In other words, the questions students posed reflected their understanding and learning of the contents.

Table 2: Correlation among Variables (N=72)

<table>
<thead>
<tr>
<th>Variable</th>
<th>Achievement (Unit1 &amp; 2)</th>
<th>Achievement (Unit3 &amp; 4)</th>
<th>Achievement (Unit5)</th>
<th>Overall Achievement</th>
<th>Mean (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>QGP at w1and 2</td>
<td>.33** ( (p=0.005) )</td>
<td>.27* ( (p=0.02) )</td>
<td>.21 ( (p=0.08) )</td>
<td>.28* ( (p=0.02) )</td>
<td>22.89 (5.12)</td>
</tr>
<tr>
<td>QGP at w3and 4</td>
<td>.21 ( (p=0.08) )</td>
<td>.29* ( (p=0.01) )</td>
<td>.17 ( (p=0.15) )</td>
<td>.23* ( (p=0.05) )</td>
<td>27.58 (6.90)</td>
</tr>
<tr>
<td>QGP at w5</td>
<td>.12 ( (p=0.34) )</td>
<td>.15 ( (p=0.22) )</td>
<td>.13 ( (p=0.29) )</td>
<td>.14 ( (p=0.26) )</td>
<td>11.26 (2.74)</td>
</tr>
<tr>
<td>Overall QGP</td>
<td>.27* ( (p=0.02) )</td>
<td>.31* ( (p=0.01) )</td>
<td>.20 ( (p=0.09) )</td>
<td>.27* ( (p=0.02) )</td>
<td>52.72 (11.61)</td>
</tr>
</tbody>
</table>

Mean (SD) | 67.50 (21.25) | 71.97 (16.58) | 72.39 (20.95) | 70.62 (18.73) |

Note: a. QGP refers to students’ question-generation performance

b. * denotes p< 0.05, ** denotes p<0.01
3.3 Conclusion

This study contributed to the literature on student question-generation. First, this study validated the effects of SQG on students reported use of cognitive strategies and metacognitive strategies. Second, this study also validated the significant correlations between students’ question—generation performance and academic achievement. The instructors who are interested in SQG strategies are suggested to teach students question-generation skills by providing them with guidance, deliberated practice opportunities and in-time feedback on their question-generation performance. As this study adopted the single group experimental design, future research is suggested to take a qualitative approach to explore the difficulty students might encounter during the question—generation process. It might help to understand any potential variables that might moderate the strength of relationship between the SQG and the three examined variables. Furthermore, this study focused on exploring the potential effects of one soft technology, the instructional strategy of SQG. As a result of the limited classroom facilities and teaching time, the hard technology, such as the online question-generation system, was not allowed to integrate in the SQG process. Future research may further explore whether adoption of the online question-posing system will further enhance students’ engagement in the SQG process.

Acknowledgements

This paper was funded by the research grant from the National Science Council, Taiwan, ROC (NSC 101-2511-S-032-006). The author would like to thank research assistants, Hao jie Yong for his assistance on data collection.

References


Structured Explanation Generation for Conceptual Understanding in Physics

Tomoya HORIGUCHI*, Takahito TOUMOTO & Tsukasa HIRASHIMA

* Graduate School of Maritime Sciences, Kobe University, Japan
b Faculty of Engineering Division II, Tokyo University of Science, Japan
c Department of Information Engineering, Hiroshima University, Japan

* horiguti@maritime.kobe-u.ac.jp

Abstract: In science education, usual problem practice hardly helps students reach ‘conceptual understanding’ with which they can solve various problems by making appropriate models of target systems. Students often superficially read the solution of a problem and apply it wrongly to others without understanding the model. It is difficult to teach how to make appropriate models because model-making expertise includes a lot of implicit knowledge. In this paper, we propose a general framework for systematically describe such knowledge, which makes it possible not only to explain various model and difference between them but also to design/sequence a set of problems appropriate for promoting conceptual understanding. Our framework was proved useful through a preliminary experiment in which the explanations generated based on our framework promoted subjects’ conceptual understanding in mechanics.

Keywords: science education, problem practice, conceptual understanding, explanation generation, semantics of constraints

1. Introduction

In science education, one serious drawback of current ‘problem practice’ is that most students fails to acquire the ability to make an appropriate model for a given task. A domain expert (such as physicist) can model not only the behavior of a system in question, but also she/he can do so in various tasks. Her/his model is always necessary and sufficient for answering the question. Such expertise consists of identifying the structure/state of the system in question and deciding the applicable principles/laws for modeling the behavior of the system. We call such ability ‘conceptual understanding’ of the domain.

Needless to say, it is very difficult for students to reach such an understanding through problem practice. Instead of considering the model, they often rely on the ‘solution’ they previously learned (e.g., the procedure of calculation). As a result, they wrongly apply the solution of one problem to another in which the solution is inapplicable. They also can't apply the solution they previously learned to another in which the solution is applicable. Even experience in many problems doesn't help them (Bransford, Brown and Cocking, 2000; VanLehn, 1998; VanLehn and van de Sande, 2009). Without models, the students occasionally succeed (by accident), and fail in many cases.

The major problem is that novice students tend to generate naive representation of a problem focusing on its superficial features (called ‘surface structure’). They can't generate the representation based on the structural features (called ‘physical structure’) (Chi, Feltovich and Glaser, 1981; Larkin, 1983; Larkin, 1985). Therefore, instead of applying principles/laws to make the model, they often apply inappropriate solution based on the superficial similarity between problems (VanLehn, 1998), or use general strategy for operating mathematical equations without considering their physical meanings (Larkin, 1981).

In order to reach conceptual understanding, therefore, students need to learn (1) to infer the structural features of problems from the superficial features, and (2) to apply appropriate principles/laws to structural features to make models necessary for solving problems. For assisting them in problem practice, it is necessary to explain not only how each problem is solved but also why the solution is possible and what physical meaning it has. That is, it must be explicit why the principles/laws are applicable to the given situation (i.e., surface structure) and what physical meaning
(physical structure) they imply. Additionally, it is important to explain not only the solution of a problem but also the relation (difference) between problems, that is, how the solution (applicable principles/laws) changes when the situation (problem) is changed. Furthermore, it would promote such learning to provide students with an appropriately designed and sequenced set of problems (Scheiter and Gerjets, 2002; Scheiter and Gerjets, 2003; VanLehn and van de Sande, 2009).

In current problem practice, such instruction has been rarely focused on, at most given by a few (experienced) teachers individually and implicitly. Especially, there have been few intelligent tutoring systems which can explain the relation between arbitrary two problems, and adaptively sequence problems considering the learning effect of order. We think this is because most of the knowledge necessary for such instruction is implicit and difficult to systematize, therefore there have been no general framework for indexing various types of problems.

In this paper, we propose a general framework for indexing problems, based on which explanation generation and problem sequencing mentioned above can be automatized. In our framework, making a model in physics is regarded as a process in which various constraints (applied principles/laws and modeling assumptions) are imposed on the target system and its behavior. A model is regarded as the set of constraints. We first formulate the model-making process in physics, then analyze the constraints which compose a model to systematize them based on their physical meanings and roles (functions). After that, we describe the applicable conditions of principles/laws in physics as a set of constraints. The constraints classified/defined in this manner are easily assigned to the situation of a problem. There are also some groups of constraints which are 'exclusive' each other (i.e., can't be valid simultaneously). Therefore, based on such classification and exclusiveness of constraints, it becomes possible to explain what physical meaning (structural features) superficial features of a problem have, what principles/laws are applicable to them and how applicable principles/laws change when the situation is changed. By indexing problems with this framework (we call it 'Semantics of Constraints: SOC'), it becomes possible to automatically extract the 'differences between problems' which is necessary for the comparison and sequencing of problems.

We first discuss the required knowledge and assistance necessary for conceptual understanding based on current research, then introduce the SOC framework. After that, we show the method for generating SOC-based explanations. The results of preliminary experiment are described which proved the usefulness of our framework. Finally, we conclude this paper and mention our future work.

2. Conceptual Understanding and Assistance

Research on problem-solving has revealed the knowledge structure domain experts in science have (Chi, Feltovich and Glaser, 1981; Larkin, 1981; Larkin, 1983; Larkin, 1985; VanLehn, 1998; VanLehn and van de Sande, 2009). Experts can (1) infer the structural features of problems with scientific concepts from the superficial features and generate the representation to which formal operations are applicable. They can also (2) generate an appropriate plan for solving the problem by operating the representation with the knowledge about qualitatively interpreted principles/laws. It is supposed that experts have acquired such knowledge by inducing the essential features through comparison of many problems and by transforming them into (some kinds of) 'schemata' or 'production rules' (VanLehn and van de Sande, 2009). It is, however, difficult for students reach such an understanding through usual problem practice. Even instructional innovations based on recent learning science research have limitedly improved students' understanding (Bransford, Brown and Cocking, 2000).

In order to promote such knowledge acquisition, it is effective to appropriately design a set of problems which includes positive/negative examples and ‘near misses’ of various problem categories and to provide them in appropriate order to students (VanLehn and van de Sande, 2009) (in fact, it is reported problem order greatly influences learning (Scheiter and Gerjets, 2002; Scheiter and Gerjets, 2003)). In order to do that, it is necessary to explicitly describe (1) the superficial/structural features of problems and their relations, and (2) qualitative interpretations of principles/laws and their means of application. However, since most of such knowledge is implicit, there have been no general framework for systematically describe such knowledge. We think this is the reason though knowledge structure necessary for expertise was revealed and an effective instructional method was proposed, it haven't been widely practiced. The framework we propose makes it possible to systematically
describe such knowledge, based on which the design of a set/sequence of problems and explanation
generation for promoting conceptual understanding become possible.

3. Semantics of Constraints

Given a physics problem (which consists of a physical system and query), one makes a model
necessary and sufficient for answering the query by embodying an appropriate part of the domain
theory. Domain theory consists of a set of propositions each of which describes a principle/law, its
applicable condition and resulting constraint(s) on the attribute(s) of the system. Constraints by
embodied principles/laws are called the 'physical phenomenon constraints (PPCs).'

In making a model, various modeling assumptions are set for selecting appropriate
principles/laws. Modeling assumptions define the structure/behavioral range of a system and physical
phenomena to be considered. Since embodied physical phenomenon constraints are valid under some
modeling assumptions, applicable conditions of principles/laws can be described with a set of
modeling assumptions. That is, a physical phenomenon constraint always has its corresponding
modeling assumptions. Constraints by modeling assumptions are called the 'modeling assumption
constraints (MACs).'

Boundary condition of a system is given by the 'boundary condition constraints (BCCs).' They
define the influence from the outside of the system. Making the influence which cannot be or need not
be calculated with a model means defining the boundary of the model (i.e., what physical processes
are considered/ignored). That is, a BCC always has its corresponding modeling assumptions.

In our framework, a model is the union of physical phenomenon constraints, boundary
condition constraints and modeling assumption constraints. Usually, only the first two constraints are
written as a model while the last constraints are remained implicit. However, MACs gives the validity
to PPCs and BCCs. When modeling assumptions are changed, physical phenomena and boundary
conditions also qualitatively change. In order to make a model correctly, therefore, it is necessary to
understand the physical meaning of the constraints based on modeling assumptions (i.e., why an
assumption is set and what role it plays). In most cases, such knowledge is acquired by a few students.
In this research, we develop a framework for describing such knowledge explicitly, based on which
the function for promoting conceptual understanding is designed. In the following two subsections,
we elaborate on each class of constraints (BCC is omitted owing to limited space) to systematize their
physical meanings and relations.

3.1 Modeling Assumption Constraints (MACs)

Modeling assumption constraints define the physical processes considered/ignored in a model. They
are classified in two ways from different viewpoints: structural and functional.

The structural viewpoint focuses on defining the structure and its state of a model. The
'physical structure constraint' specifies what kind of objects, relations and their attributes in a system
are considered. It corresponds to selecting a viewpoint, granularity or coordinate system of a system.
An example is the specification about whether their mechanical relations/attributes (e.g., mass,
applied forces) or their electrical ones (e.g., current, resistance) are considered. On the other hand, the
'operating range constraint' specifies the range within which a model is valid since physical
phenomena occur assuming a system is in a specific state. For example, a model of a resistance
assuming its value is constant needs the specification that its current and voltage are within the
proportional range.

The functional viewpoint focuses on defining the boundary of a model to specify what kind of
physical processes are considered/ignored. The 'process consideration constraint' makes such selection
about physical processes of the same granularity (where, the 'out-sourcing/black-boxing constraint'
ignores a physical process by put-ting it out of the system or into a black box regarding its effect as a
boundary condition, and the 'process selection constraint' simply ignores a physical process and its
effect). For example, assuming constant voltage supplied from outside is an out-sourcing constraint.
Considering two parallel-connected resistors as a compound re-sister is a black-boxing constraint.
Considering/ignoring the friction between two objects is a process selection constraint. The 'physical
world constraint' maintains the fundamental laws of the physical world, such as 'rigid objects never
overlap.' More microscopic physics is necessary to explain why this constraint is valid, that is, it
specifies the physical processes of smaller granularity are ignored. The 'process simplifying
constraint' substitutes the simplified process for an original complicated process in order to make the

361
(mainly mathematical) calculation with a model easier. An example is to consider the behavior of a pendulum with small amplitude as simple harmonic oscillation (not as circular motion).

Constraint classes from the structural viewpoint are useful for enumerating modeling assumptions because they rather suggest the components and their relations of a system. For example, when a variable in an equation stands for a physical quantity, it is easy to infer an object and its attribute corresponding to the quantity is considered (which are physical structure constraints). Constraint classes from the functional viewpoint are useful for considering the meaning of modeling assumptions because they rather suggest the process structure (processes considered and their relations). For example, considering/ignoring a physical attribute (which is a physical structure constraint) suggests a physical process concerning the attribute is considered/ignored (which is a process selection constraint). That is, the classes from the structural view-point rather concern the surface structure of a problem, while the classes from the functional viewpoint rather concern its physical structure. Furthermore, as shown above, the classes from both viewpoints are related to each other based on their physical meanings. Therefore, with these classifications, it becomes possible to systematically describe the knowledge about the relation between superficial and structural features of problems.

Additionally, there are often sets of modeling assumption constraints which are mutually exclusive (can't be assumed simultaneously). For example, in the same time interval, 'transient state' and 'steady state' (which are operating range constraints) can't be assumed simultaneously. In the same (part of a) system, 'consider friction' and 'not consider friction' (which are process consideration constraints) can't be assumed simultaneously. Such exclusiveness between modeling assumptions gives important clues to identify the differences between models/problems (see section 4.2).

### 3.2 Physical Phenomenon Constraints (PPCs)

Relatively simpler physical phenomenon constraint is the 'physical device constraint' which arises within a component of a system. That is, it is a 'local constraint.' Since it indicates the physical property of the component, each domain has its specific physical device constraints (for example, Ohm's law constrains the values of current and voltage in an electric device). In contrast, there are 'global constraints' which indicates the behavior of multiple components or the whole system. Global constraints are classified as follows.

In general, a physical system evolves through time, starting from an initial state. It is either (1) changing dynamically, (2) in a steady state or (3) changes discontinuously. Therefore, we call the constraints in these states, the 'dynamic change constraint,' the 'steady state constraint' and the 'discontinuous change constraint,' respectively. Additionally, when a quantity is conserved through time, the constraint which indicates its amount is the same at arbitrary two time points is called 'conservation law constraint.'

Dynamic change constraint constrains the behavior of a system in a time interval during which it is changing dynamically. It often indicates the relation between the driving power of dynamic change and the influences on it. For example, Newton's second law (equation of motion) relates an object's acceleration with the forces applied to it. Steady state constraint constrains the behavior of a system in a time interval during which it is in a steady state. It indicates the balance/cancellation between influences on the driving power of dynamic change. An example is the equation of balance of forces about an object at rest. Discontinuous change constraint constrains the behavior of a system at a time point on which it changes discontinuously. It indicates the relation between the amounts of a quantity before and after the change. An example is the formula of coefficient of restitution. A quantity is called a 'conserved quantity' when its amount is constant during the temporal evolution of a system. Conservation law constraint indicates the amounts of a conserved quantity at arbitrary two time points are the same. An equation of heat exchange between two objects and an equation of conservation of energy/momentum are the examples.

A global physical phenomenon constraint aggregates a set of local physical phenomenon constraints. For example, Newton's second law (equation of motion), which is a dynamic change constraint in mechanics, includes a set of local constraints each of which indicates a force applied to the target object (physical device constraints such as elastic force, friction). Such inclusion relation between PPCs gives important clues to identify the dominant principle(s)/law(s) in solving a problem.

Additionally, there are often sets of physical phenomenon constraints of which modeling assumptions (preconditions) are mutually exclusive. These PPCs are never simultaneously valid in the same state of the same system. For example, since 'static friction' and 'kinetic friction' have exclusive
preconditions (operating range constraints) about a contact surface of two objects, they are never valid simultaneously at the same surface. The first three global PPCs (i.e., dynamic change, steady state and discontinuous change constraints) are exclusive for the same reason. They often entirely change each other when preconditions are changed. For example, suppose a mechanical system is in a steady state by assuming 'friction' which cancels other forces. When the assumption is changed to 'frictionless,' the system can dynamically change. Such exclusiveness between PPCs gives important clues to identify the differences between models/problems (see section 4.2).

4. Explanation Generation

4.1 Framework of model-making process description

In our framework, each principle/law is described as a 'model fragment' (Falkenhainer and Forbus, 1991) which consists of its applicable condition and its consequence(s). Applicable condition is described as a set of modeling assumption constraints, while a consequence is described as a physical phenomenon constraint. A model consists of the union of PPCs given by instantiated model fragments, MACs giving applicable conditions for them, and boundary condition constraints given in a problem. (Note that an instantiated 'model fragment' is distinguished from 'model fragment class' which describes a principle/law itself.) A model-making process (i.e., solution) is described as the procedure in which model fragments are applied (instantiated) in turn to the situations (represented with MACs and BCCs) to yield new consequences (represented with PPCs). (Note that a consequence of a model fragment can be the condition for others.)

Figure 1a and 1b show examples, in which it is explicitly described why/how each principle/law is applied to the given situation. In contrast, usual description of solution focuses on the calculation of the required physical amount from the given ones, while the principles/laws and conditions which justify the calculation are attached in the ad hoc way. SOC enables implicit assumptions and physical meanings of calculation to be systematically described.

Additionally, a pair/set of model fragment classes which have similar conditions (situations) but have exclusive MAC(s)/PPC(s) as applicable condition(s)/consequence(s) is called 'exclusive model fragment classes.' Grouping such model fragment classes helps the comparison of models.

4.2 Procedure

4.2.1 Explanation of the model-making process (solution)

The description of model-making process mentioned above makes it possible to explain why/how each principle/law is applied explicitly referring to its modeling assumptions. In figure 1b, for example, the formula \(v^2 - v_0^2 = 2ax\) (dynamic change constraint) is used which relates an object's displacement, velocity and acceleration in a time interval. Note that the constraint 'acceleration is constant in the interval' (operating range constraint) is explicitly described which is an important precondition for this model fragment to be applied. Many students wrongly use this formula when an object's acceleration temporally varies. The explanation explicitly referring to modeling assumptions would be helpful in avoiding such mistakes.

Additionally, in solving problems, it is important to recognize not only each local principle/law and its consequence, but also the global principle/law which dominates the behavior of the whole system. The solution of domain experts is often 'dominant-principle/law-driven,' that is, they first recognize the dominant principle/law of a problem, then apply local principles/laws to 'fill in the slots' of the global principle/law (Chi, Feltovich and Glaser, 1981; Larkin, 1983; VanLehn and van de Sande, 2009). In our framework, a model fragment of global physical phenomenon constraint (PPC) are defined as the aggregation of the model fragments of local PPC which compose the global one (the applicable condition of a global model fragment is the union of its component model fragments). Global model fragments make it possible to explain the model-making process (solution) focusing on the dominant principle/law. For example, in figure 1a, the model fragment 'balance-of-forces' gives a steady state constraint (global PPC) and its applicable condition includes some physical device constraints (local constraints) given by other model fragments. Based on such inclusion relation, the sequence of explanation can be controlled as follows: first, to indicate the given condition 'a block is
at rest' (which means its velocity doesn't temporally vary) suggests 'balance of forces' should be used, then to refer to the laws 'gravity' and 'static friction' which influence the driving power of the block's velocity. The generated explanation is shown in figure 2a.

Problem:
(a) What is the minimum value of $\mu_s$ in order for the block to be rest?
(b) Derive the velocity of the block when it arrives at the bottom of the slope.

(a) Q: What is the minimum value of $\mu_s$ in order for the block to be rest?

(b) Q: Derive the velocity of the block when it arrives at the bottom of the slope.

Figure 1. Examples of model-making process.

(a) Since an object keeps at rest, the principle 'balance of forces' is used. So, all the forces applied to the object are required (in the parallel direction to the include plane). Gravity $mg$ is applied downwards and its component parallel to the plane is $mgsin\theta$. Since the plane's surface is rough, static friction $\mu mgcos\theta$ is applied. These forces balances with each other, that is, $mgsin\theta = \mu mgcos\theta$. From this equation, $\mu$ must be greater than or equal to $tan\theta$. (partly omitted)

(b) Since an object moves at uniform acceleration, the principle 'Newton's second law' is used. So, all the forces applied to the object are required (in the parallel direction to the include plane). Gravity $mg$ is applied downwards and its component parallel to the plane is $mgsin\theta$. Since the plane's surface is rough, kinetic friction $\mu kmgcos\theta$ is applied. These forces influences the acceleration of the object, that is, $ma = mgsin\theta - \mu kmgcos\theta$. From this equation, $a = g(sin\theta - \mu kcos\theta)$. Since the acceleration is constant, the formula $v^2 - v'^2 = 2ax'$ is applicable. Consequently, the velocity of the object at the bottom of the slope is $\sqrt{2lg(sin\theta - \mu kcos\theta)}$. (partly omitted)

Figure 2. Examples of generated explanation.

4.2.2 Explanation of the difference between models (problems)

The difference between models (problems) can be inferred by comparing their model fragments. There are two types of relations between problems: (1) the problems which have the same/similar surface structures (situations) but have different physical structures (instantiated model fragments belong to different classes) and (2) those which have different surface structures (situations) but have the same/similar physical structures (instantiated model fragments belong to the same classes).
relations play an important role for promoting conceptual understanding (Scheiter and Gerjets, 2002; Scheiter and Gerjets, 2003). As for the latter, the difference is easily inferred by identifying the corresponding pair of model fragments (each of which belongs to each model) both of which give the (global) PPCs of the same class. The difference can be explained by showing their preconditions (situations) are different.

As for the former, the difference is inferred by identifying the corresponding pair of model fragments (each of which belongs to each model) which belong to exclusive model fragment classes. Since their situations are similar but their modeling assumption constraint(s) and physical phenomenon constraint(s) are exclusively different, they indicate the difference of two models before/after the change of the situation. The type of the difference can be explained by referring to their modeling assumption classes. For example, when two corresponding model fragments have the same physical structure constraints and exclusively different operating range constraints, it is inferred that the difference of two models is change of the operating range about the partial system they match. The model fragments 'static-friction' in figure 1a and 'kinetic-friction' in 1b are in such relation. It can be inferred that the local constraint between a block and slope is changed from 'static-friction' to 'kinetic-friction' by changing the operating range, by which the global constraint 'balance of forces' (steady state constraint) is changed to 'Newton's second law' (dynamic change constraint) (the generated explanation is shown in figure 2b).

Additionally, when comparing models (problems), it is important to recognize not only the change of each local principle/law and its consequence, but also the change of the global principle/law which dominates the behavior of the whole system. Global model fragments which aggregate the model fragments of local PPCs, make it possible to explain the behavioral change of the whole system focusing on the dominant principle/law.

5. Preliminary Experiment

5.1 Design

We conducted an experiment to evaluate the usefulness of our framework. A SOC-based explanation generator was implemented. The purpose was to examine whether the SOC-based explanation promotes students' conceptual understanding, that is, whether their representation of problems was improved and they became able to solve various types of problems by using correct models.

Subjects: Fifteen graduates and under graduates whose majors are engineering participated in.

Instruments: (1) Two sets of problems in elementary mechanics: They were called 'problem set 1 (PS-1)' and 'problem set 2 (PS-2).’ Each set included fifteen problems of various surface/physical structures. Problems might have similar situations but different solutions, or have different situations but similar solutions. The sets had no common problem. (2) Usual explanation about the solutions of eleven problems in PS-1: The calculation of the required physical amount from the given ones was mainly explained. (3) SOC-based explanation about the solutions of the same problems as usual explanation: In addition to the solution of each problem, the differences between problems were explained about eight pairs of problems which had similar surface/physical structures. (4) Explanation generator used for generating SOC-based explanation: Model-making processes described by the experimenter (first author) were input and their explanations were output, which were rewritten into readable natural language by the experimenter (without changing the point).

Procedure: First, subjects were given PS-1 and asked to group the problems into some categories based on some kind of 'similarity' they suppose (any number/size of categories were allowed), then asked to label each category they made (called 'categorization task 1'). After that, they were asked to solve eight problems in PS-1 (called 'pre-test'). After a week, the subjects were divided into two groups: one was the 'control group' (seven subjects) and another was the 'experimental group' (eight subjects). The average scores of both groups in pre-test were made equivalent. The subjects in control group were given the usual explanation and asked to learn it. The subjects in experimental group were given the SOC-based explanation and asked to learn it. After that, by using PS-2, 'categorization task 2' was conducted in the same way as above. Finally, subjects were asked to solve eight problems in PS-2 (called 'post-test').

Measure: The quality of the representation of problems was measured with the categories, their 'frequencies' (number of problems accounted for) and the time required in each categorization task. The ability to solve various types of problems was measured with the scores in each test.
learning with usual/SOC-based explanation on the quality of representation and the ability of problem-solving was measured with the comparison of the results of two categorization tasks and pre-/post-tests. The superiority of SOC-based explanation to usual explanation was measured with the differences of improvement of categorization and problem-solving between experimental and control groups.

5.2 Results

The categories made by subjects and their frequencies in categorization task 1 are shown in table 1. Most of the subjects categorized the problems based on the similarity of their superficial features, such as the components of the system (e.g., inclined plane, springs), the figures of motion (e.g., circular motion, free fall). Additionally, all subjects finished the task within ten minutes. The results of categorization task 2 are shown in table 2 (for control group) and table 3 (for experimental group). Many subjects of control group still categorized the problems based on the similarity of their superficial features, while many subjects of experimental group became to categorize the problems based on the similarity of their structural features, that is, the dominant principles/laws of problems (e.g., Newton's second law, balance of forces, conservation of energy). Additionally, all subjects of control group finished the task within ten minutes again, while the subjects of experimental group required from twenty-five to thirty-five minutes. These results suggest that the learning with SOC-based explanation promoted representing problems based on their structural features rather than their superficial features (the increase of the time required suggests the subjects of experimental group inferred the physical structure from surface structure).

The average scores in pre- and post-tests are shown in figure 3 (in both tests, full marks were 52). In pre-test, there was no significant difference of average scores between groups (control group: 36.0 and experimental group: 33.6, t-test p >.10). In post-test, though there was also no significant difference of average scores between groups (control group: 42.7 and experimental group: 47.6, t-test: p > .10), the increase of average score of experimental group was larger than that of control group. This result suggests that the learning with SOC-based explanation promoted the ability to solve various types of problems, that is, to make appropriate models regardless of their superficial features.

These results suggest that SOC-based explanation about the solution of problems and their differences can assist students in reaching conceptual understanding.

<table>
<thead>
<tr>
<th>Table 1: Categories in task-1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of subjects using category labels (N1=15)</td>
</tr>
<tr>
<td>---</td>
</tr>
<tr>
<td>Springs</td>
</tr>
<tr>
<td>Free fall etc.</td>
</tr>
<tr>
<td>Collision</td>
</tr>
<tr>
<td>Circular motion</td>
</tr>
<tr>
<td>Acceleration</td>
</tr>
<tr>
<td>Strings</td>
</tr>
<tr>
<td>Inclined planes</td>
</tr>
<tr>
<td>Balance</td>
</tr>
<tr>
<td>Object only</td>
</tr>
<tr>
<td>Friction</td>
</tr>
<tr>
<td>Second law</td>
</tr>
<tr>
<td>Pulleys</td>
</tr>
<tr>
<td>Balance of energies</td>
</tr>
<tr>
<td>Motion of weight</td>
</tr>
</tbody>
</table>
Table 2: Categories in task-2 (usual)

<table>
<thead>
<tr>
<th>Category</th>
<th>Number of subjects using category labels (N₁=7)</th>
<th>Average size of category (N₂=15)</th>
<th>Number of problems accounted for (N=N₁×N₂=105)</th>
<th>Number of problems wrongly accounted for (N*=105)</th>
<th>Number of problems correctly accounted for (N=105)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Springs</td>
<td>4</td>
<td>4.5</td>
<td>18</td>
<td>0</td>
<td>18</td>
</tr>
<tr>
<td>Inclined planes</td>
<td>4</td>
<td>3.3</td>
<td>13</td>
<td>0</td>
<td>13</td>
</tr>
<tr>
<td>Balance of forces</td>
<td>3</td>
<td>3.7</td>
<td>11</td>
<td>0</td>
<td>11</td>
</tr>
<tr>
<td>Conservation of energy</td>
<td>3</td>
<td>6.0</td>
<td>18</td>
<td>9</td>
<td>9</td>
</tr>
<tr>
<td>Second law</td>
<td>3</td>
<td>3.7</td>
<td>11</td>
<td>2</td>
<td>9</td>
</tr>
<tr>
<td>Pulley and string</td>
<td>2</td>
<td>3.5</td>
<td>7</td>
<td>0</td>
<td>7</td>
</tr>
<tr>
<td>Circular motion</td>
<td>4</td>
<td>1.5</td>
<td>6</td>
<td>0</td>
<td>6</td>
</tr>
<tr>
<td>Pendulum</td>
<td>3</td>
<td>1.7</td>
<td>5</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>Simple harmonic motion</td>
<td>2</td>
<td>2.0</td>
<td>4</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>Collision</td>
<td>2</td>
<td>1.0</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
</tbody>
</table>

Table 3: Categories in task-2 (SOC)

<table>
<thead>
<tr>
<th>Category</th>
<th>Number of subjects using category labels (N₁=8)</th>
<th>Average size of category (N₂=15)</th>
<th>Number of problems accounted for (N=N₁×N₂=120)</th>
<th>Number of problems wrongly accounted for (N*=120)</th>
<th>Number of problems correctly accounted for (N=120)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Balance of forces</td>
<td>7</td>
<td>4.4</td>
<td>31</td>
<td>5</td>
<td>26</td>
</tr>
<tr>
<td>Second law</td>
<td>7</td>
<td>3.6</td>
<td>25</td>
<td>1</td>
<td>24</td>
</tr>
<tr>
<td>Conservation of energy</td>
<td>8</td>
<td>4.1</td>
<td>33</td>
<td>12</td>
<td>21</td>
</tr>
<tr>
<td>Linear accelerated motion</td>
<td>3</td>
<td>3.3</td>
<td>10</td>
<td>2</td>
<td>8</td>
</tr>
<tr>
<td>Conservation of momentum</td>
<td>3</td>
<td>1.3</td>
<td>4</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>Acceleration</td>
<td>1</td>
<td>3</td>
<td>3</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>Springs</td>
<td>1</td>
<td>3</td>
<td>3</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>Pulleys</td>
<td>1</td>
<td>3</td>
<td>3</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>Simple harmonic motion and period</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>String and tension</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>Time</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>2</td>
</tr>
</tbody>
</table>

367
6. Conclusion

Aiming at promoting conceptual understanding through problem practice, we proposed the SOC framework based on which the knowledge necessary for designing a set of problems, sequencing them and generating explanations can be described. We showed the explanations generated with our framework could promote conceptual understanding through a preliminary experiment. SOC-based explanation generator can provide a basic function for designing various instructional methods (e.g., a detailed explanation is gradually simplified (scaffolding-fading), a sequence of problems is given which promotes spontaneous induction). Design of such instructional methods and verification of their effectiveness are our future work.

References

Practical Use of Interactive Environment for Learning by Problem-posing for One-step Multiplication and Division Word Problems

Sho YAMAMOTO*, Yuki AKAO*, Mitsutaka MUROTSU*, Takehiro KANBE*, Yuta YOSHIDA*, Kazushige MAEDA**, Yusuke HAYASHI* & Tsukasa HIRASHIMA*

*Graduate School of Engineering, Hiroshima University, Japan
**Elementary School Attached to Hiroshima University, Japan
*sho@lel.hiroshima-u.ac.jp

Abstract: Problem-posing is known that effective method for promoting to master the use of solution methods. However, these methods have been rarely used because of the activity for posing problem and assessment the posed problem. So, we design and develop an interactive environment for learning by problem-posing continually. In previous research, we have already designed and developed the learning environment targeting one-step addition or subtraction word problems and one-step multiplication word problems. In this paper, we have designed and developed the learning environment targeting one-step multiplication or division word problem and its assignment newly. For realizing this system, the property of quantity and its relation is suggested as problem structure. And the diagnosis and feedback of posed problem are defined based on the property and relation. The level of assignment is defined too. Developed learning environment and its practical use in an elementary school are reported.

Keywords: Problem-posing, sentence-integration, multiplication word problems, division word problems, problem structure

1. Introduction

Learning by problem-posing is well known as effective method for promoting learners to master the use of solution methods (Polya, 1945; NF Ellerton, 1986; Silver, CAI, 1996). Moreover, it has been proposed that poor problem solvers often fail to elicit problem structures from problem (Brown, VanLehn, 1980; Mayer, 1982; Kintsch, Greeno, 1985). So, it is postulate that learning by problem-posing is effective for learner to promote to acquire the problem structure. However, this exercise also known that it is difficult to perform because of the cost of activity for posing problem and assessment the posed problem. Therefore, we design and develop the learning environment which learners acquiring the structure of arithmetic word problem by exercising the problem-posing continually (Nakano, et al, 1999; Hirashima, et al, 2007; Hirashima, et al, 2011). This learning environment is required a learner to pose a problem by selecting three cards from a set of given sentence cards and arranging them in proper order (Hirashima, et al, 2014). Also, the learning environment can generate feedback about posed problem. We call this learning environment as "MONSAKUN".

Until now, one-step addition or subtraction word problems and multiplication word problems are analyzed, and the structures of these problems are implemented on tablet PC for realizing the problem-posing exercise and an assessment of posed problem (Yamamoto, et al, 2012; Yamamoto, et al, 2013). MONSAKUN consists of MONSAKUN Touch and MONSAKUN Analyzer. By using this environment, a learner can exercise the problem-posing on MONSAKUN Touch, and a teacher can confirm the result of learner's problem-posing on MONSAKUN Analyzer via network. For this implementation, teacher can use our learning environment in their arithmetic class and lecture the arithmetic word problem by the problem-posing. Actually, in addition to the development of the learning environment, we have performed two experimental uses with elementary school teacher.
First experimental use is, which first grade students were used by MONSAKUN Touch for posing problem that can be solved by one-step addition or subtraction word problems because first grade students have just learned one-step addition or subtraction in arithmetic class. In second experimental use, MONSAKUN Touch 2 is used by the second grade students, which for posing problem that can be solved by one-step multiplication word problems because second grade students have just learned one-step multiplication in arithmetic class. The results of these experimental uses have proposed that not only the learner improve the problem solving performance, but also this learning environment was effective for the learner who can’t judge the problem structure to acquire it.

Based on these researches, we have designed and developed a learning environment for posing problems that can be solved by one-step multiplication or division. Since third grade students learn not only one-step multiplication word problems but also one-step division word problems in contrast to second grade students, these scopes are targeted in our research as next step. In order to realize this, the design of a model of problem-posing and an assignment of problem-posing for the learning environment based on analysis of targeted problem should be performed. In this paper, a problem structure is explained in the following chapter. A design of developed learning environment based on this structure is described in section 3. A sequence of assignment is also expressed. Subsequently, a procedure of its practical use and an analysis of the results are reported.

2. Problem Structure of One-step Multiplication or Division Word Problem

In this section, the model of one-step multiplication or division arithmetic word problem is explained. One-step word problems can be expressed by three sentences in our research. Example is shown in Figure 1. Because there are three values in one-step arithmetic word problem, this problem can be expressed by three sentences. These sentences consist of two sentences mean existence and one sentence means relation between other two values. We call each sentence as existence sentence and relation sentence. In this example, “There are three boxes” and “There are several apples” are existence sentence. “There are four apples in each box” is relation sentence because this sentence shows the relation between the apple and box. These sentences consist of value, object and predicate.

In addition to the kind of sentence, multiplication and division word problem have a property of quantity (Yamamoto, et al, 2013). Generally, multiplication is expressed by “multiplicand multiplied by multiplier is product” (Greer, 1992; Vergnaud, 1983; Davies, 1841). In other words, each quantity has different property. In Figure 1, multiplicand is expressed as "There are five apples in each box", multiplier is "There are three boxes" and product is "There are several apples". That problem contains the story that the value of apples is expressed as the amount of apple when there are three boxes and the value of apples in each box is basis. Since, in Japanese Education, multiplicand is also called "base quantity", multiplier is "proportion" and product is "compared quantity". Then, the arithmetic word problems that can be solved by one-step multiplication or division has three types of story. This is, (1) Compared quantity divided by base quantity is proportion, (2) Base quantity multiplied by proportion is compared quantity, (3) Compared quantity divided by proportion is base quantity. The story of the problem in Figure 1 is (2).

All of these stories contain the relation that is “Base quantity multiplied by proportion is compared quantity”. One-step multiplication or division word problems are expressed by changing the one quantity to required value in each story. Therefore, it is important to extract the base quantity, proportion and compared quantity from problem and to make the relation between these quantities with “Base quantity multiplied by proportion is compared quantity”. Next section, the implementation of the problem structure mentioned above to tablet PC is described.
3. Learning Environment for Problem-posing "MONSAKUN Touch3"

3.1 Framework of Learning Environment

This learning environment consists of MONSAKUN Touch 3 for learners and MONSAKUN Analyzer 3 for teachers. A result of the learner’s learning by problem-posing on MONSAKUN Touch 3 is sent to database server via network. This framework is shown in Figure 2. MONSAKUN Touch 3 developed by using Android, MONSAKUN Analyzer 3 by using PHP and JavaScript. The each software can be run on Android Tablet. RDBMS is used MySQL. The teacher can confirm the graph of learner’s learning by using MONSAKUN Analyzer 3 that receives a learning data from database server. The learning data are saved as three data: the number of correct problem, the number of incorrect problem, the number of the each incorrectness and the learner’s log. Category of incorrectness is explained in next section. MONSAKUN Analyzer 3 generates some graphs by using these data and displays teacher it. Teacher can limit to an assignment that learners can exercise on MONSAKUN Touch 3 by using MONSAKUN Analyzer 3.

3.2 MONSAKUN Touch 3

3.2.1 Flow by Using MONSAKUN Touch 3

In this section, the flow of problem-posing by using MONSAKUN Touch 3 is described. This is same as the MONSAKUN Touch 1 and 2 but an interface of problem-posing is a little changed. First, a learner logs in this system by selecting his/her grade, class and number. After that, the learner selects a level of assignment on an interface for selecting level. The learner is also able to select particular assignment and switch feedback on or off. The level of assignment is elaborated in 3.2.3. After selecting the level, he/she sees an interface of problem-posing is shown in Figure 3. This interface presents the assignment for posing problem, the set of given sentence card and three blank for arranging given sentence cards. Sentence cards are written sentence like “There are three apples".
These sentence cards were mentioned in previous section. Hence the learner can pose the problem by selecting three sentence cards from given cards and arranging them in proper order. Given sentence cards are consists of correct card set and dummy card set for leading to errors. In the MONSAKUN Touch 3, the text means the property of quantity is shown in the left side of each blank because the environment lets the learner consider the property of each sentence card when they learn by using this environment. If three blank is filled with three sentence cards, diagnosis button will be active. Then, the learner can tap this button and the system diagnoses and feed back his/her posed problem. When the learner finishes answering all assignment in selected level correctly, the interface for posing problem backs to the interface for selecting level.

\[\text{Figure 3. Main Interface of MONSAKUN Touch 3.}\]

3.2.2 Diagnosis and Dummy Card

Figure 4 is the procedure for diagnosing of posed problem. This procedure is processed based on the problem structure is described in chapter 2. The posed problem is required to satisfy these several constraint. First, MONSAKUN Touch 3 assesses the composition of sentence. This is, the combination of existence sentences and relation sentence, and the setting cards and its property of quantities. If these are not correct, MONSAKUN Touch 3 gives a learner feedback it. If these are correct, next, MONSAKUN Touch 3 assesses whether a composition of story is correct or not. This means that the system checks the relation of each object, value, unit in answered three sentence cards. In addition to this, the learning environment assesses not only the relation between two objects of base quantity and object of other sentence cards but also the base value of base quantity like “apple per 2 boxes”. After that, MONSAKUN Touch 3 assesses whether the calculation expression of posed problem and given calculation by assignment is same or not. If learner causes the error, the system feedbacks its reason based on diagnosis in Figure 4. When posed problem satisfies the above constraint, this problem is diagnosed as correct.

Dummy card is included in given sentence card in order to let the learner cause the error that is show in Figure 4. These cards make by changing object, number or predicate of correct card. Because of dummy cards, the learner needs to be vividly aware of problem structure when they posed problem on MONSAKUN Touch 3.
3.2.3 Designing the Level of Assignment

In this research, we have designed the level of assignment gradually so that the learner acquires the problem structure seamlessly. Table 1 shows the all level of assignment by dividing into the number of level, assignment, required activity, contents of assignment and number of assignment. Then, each level designed on the basis of "Base quantity multiplied by proportion is compared quantity" for third grade students on elementary school. The learner is required to pose the story from level 2 to 7, to pose the problem from level 8 to 9.

In level 1, the learner is given the story of one-step multiplication and four calculation expression which are expressed by "Base quantity multiplied by proportion is compared quantity", "Proportion multiplied by base quantity is compared quantity" and the cumulation of same number like “4+4+4=12” and “3+3+3+3=12”. This assignment is the confirmation of the relation between multiplication and addition. Then, the learner is required to select the correct calculation expression. The purpose of this level is which let the learner comprehend the relation of multiplication story and addition calculation. The learning environment gives the story and several sentence cards to the learner in level 2. The given story as sentence integration consists of two fixed sentence cards and one blank. The learner is required to fill this blank by considering the property of quantity. In this assignment, they learn the property of quantity that is contained each given sentence card. Given sentence cards in level 3 are included two sentence cards that have different text representation and same property. For example, "There are two boxes." and "The number of box is two.". In this level, let the learner learn that the sentence cards include the same property of quantity have various text representation. MONSAKUN Touch 3 presents the three blank for putting the sentence cards and several sentence cards in level 4. Then, the learner is required to pose the story by selecting three sentence cards and by arranging them in proper order based on the relation of "Base quantity multiplied by proportion is compared quantity". The assignment of level 5 requires the learner to pose the two stories by using one common sentence card. For example, "There are three apples in each box. There are six boxes." and "There are two boxes in each shelf. There are three shelves. There are six boxes" are used same card that is "There are six boxes.". Through this exercise, the learner comprehends that existence sentence card is able to have two property of quantity. In other words, both proportion and compared quantity are expressed by existence sentence. After that, in level 6, the learner learns that the story has three kinds of calculations expression that are mentioned in section 2. This purpose is that the learner notices the multiplication story contains the calculation (a) and (c). Thus, the learner is given the multiplication and division calculation expression as assignment for posing story. In order to let the learner confirm three properties of quantity and its relation again, assignments of level 7 includes improper assignment which cannot solve because of lack of one proper sentence card. Then, the learner is given a specific sentence card for posing the story in this level, which is labeled "proper sentence card is not given" instead of lacking sentence card. Because the assignments in level 7 are composed of usual assignment and
assignment which mentioned above, the learner is required to consider each property of quantity and its relation again. As the next step, the learner is required to pose problem in level 8 because the learner learn to pose the story through level 2 to 7. Finally, in level 9, the learner is required to pose the two problems by using one common sentence card. This assignment is same as assignment of level 5. Through the exercise from level 1 to 9, the learner can acquire the problem structure gradually.

Table 1: The Assignment Level on MONSAKUN Touch 3.

<table>
<thead>
<tr>
<th>Level</th>
<th>Required activity</th>
<th>Contents of assignment</th>
<th>Number of assignment</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Select calculation expression</td>
<td>Select calculation express given story</td>
<td>12</td>
</tr>
<tr>
<td>2</td>
<td>Pose story</td>
<td>Pose story that is expressed by given calculation (one-step multiplication) Required story has already given two sentence cards</td>
<td>12</td>
</tr>
<tr>
<td>3</td>
<td>Pose story</td>
<td>Same as assignment of level 2 Include same property and different text representation</td>
<td>12</td>
</tr>
<tr>
<td>4</td>
<td>Pose story</td>
<td>Pose story that is expressed by given calculation (one-step multiplication) Select three sentence cards and arrange them</td>
<td>10</td>
</tr>
<tr>
<td>5</td>
<td>Pose story</td>
<td>Pose two stories by using same sentence card</td>
<td>10</td>
</tr>
<tr>
<td>6</td>
<td>Pose story</td>
<td>Pose story that is expressed by given calculation But given calculation expression is one-step multiplication or division</td>
<td>12</td>
</tr>
<tr>
<td>7</td>
<td>Pose story</td>
<td>Same as assignment of level 6 But one proper sentence card is not given</td>
<td>12</td>
</tr>
<tr>
<td>8</td>
<td>Pose problem</td>
<td>Pose problem that is expressed by given calculation Select three sentence cards and arrange them</td>
<td>12</td>
</tr>
<tr>
<td>9</td>
<td>Pose problem</td>
<td>Pose two problems by using same sentence card</td>
<td>12</td>
</tr>
</tbody>
</table>

3.3 MONSAKUN Analyzer 3

Here, MONSAKUN Analyzer 3 for visualizing the learner’s learning data on MONSAKUN Touch 3 is explained in line with the function of MONSAKUN Analyzer 3. After the teacher loged in the learning environment by inputting id and password, MONSAKUN Analyzer 3 changes the interface that is shown in Figure 5. In this interface, the learning environment displays the average of student’s learning data in each lesson that are received some learning data from database server. This interface generates and shows a three bar charts and a doughnut chart. Three bar charts consist of the average number of posed problem in total, the average number of correct problem and the average number of incorrect problem. A doughnut chart shows the rate of each error that category is classified in 3.2.2. In addition to this information, MONSAKUN Analyzer 3 indicates the average progress of the level and assignment number in class. These displayed learning data can be filtered out based on the each level and assignment in each class. MONSAKUN Analyzer 3 displays these graphs not only in each lesson but also in each student by clicking the link “see the each student’s data”. The interface element is same as Figure 4 but each graph are visualized in each student’s learning data in total. This interface also only displays the learning data in each level or each assignment. Then, the teacher can see the posed problem of each student by clicking "see the each student's log". These data are updated in real-time. The teacher is able to arrange his/her lesson on the basis of these visualized data. For example, by seeing the learning data in each student, the teacher can know the students who failure in the exercise of MONSAKUN Touch 3 and support their exercise.
4. Experimental Use of MONSAKUN Touch 3

4.1 Procedure of Experimental Use

Subjects were thirty-nine students in the third grade of an elementary school. They were divided into subjects who experienced MONSAKUN and who did not experience it in previous experimental use of our research (Yamamoto, et al, 2012; Yamamoto, et al, 2013). Inexperienced group of MONSAKUN has learned one-step addition, subtraction or multiplication word problem by usual lesson only. Moreover, they had just learned to solve arithmetic word problems that can be solved by one-step multiplication or division. This experimental use has been performed during thirteen lessons that consist of pretest in one lesson, eleven lessons by using MONSAKUN and posttest in one lesson (45 minutes per lesson, in 5 weeks). A lesson by using MONSAKUN has composed of teaching about problem-posing by a teacher and problem-posing exercise by using MONSAKUN Touch 3. The teacher decided the time of using MONSAKUN Touch 3 based on the progress of each lesson. If the subjects have finished twice the current level when they exercise the problem-posing after teaching, they were allowed to work on the previous level. The purpose of this experimental use is to examine the effects of MONSAKUN Touch 3 by using a usual problem solving test, an extraneous problem solving test and a problem-posing test, and the effects of experience MONSAKUN is also examined.

We used these three tests: the problem solving test, the extraneous problem solving test and the problem-posing test. The problem solving test is the usual problem solving test can be solved by one-step multiplication or division that is expressed by three sentences. This test is included five stories that are made as the permutation of base quantity, proportion and compared quantity without "Proportion multiplied by base quantity is compared quantity" because of commutative law. Therefore, the usual problem-posing test has fifteen questions because each quantity can be the required value in these five stories. Extraneous problem solving test includes extraneous information.
that is not necessary to solve the problem. It is more difficult for learner to solve the extraneous problem than to solve the usual problem solving (Muth, 1992). The subjects are required to judge the relevance of each sentence and find the sentence including as the extraneous information for solving the problem. Therefore, the extraneous problem solving test is useful to assess learner's comprehension of the problem structure. These problems consists of twelve problems that including the two kinds of extraneous information that change sentence cards except sentence contains required value in each six stories. The problem-posing test examines the problem-posing performance to let the subject pose the problem as he/she can within the time limit. The subject pose problem from scratch. The time limit is ten minutes in each test. The difference between pretest and posttest is order of each problem.

4.2 Analysis of Pretest and Posttest

An analysis of pretest and posttest are reported in this section. And the level by using lecture is described. The teacher has performed the lecture based on the level on MONSAKUN Touch 3 and treated one level in one lecture. However, the subjects can not relate between multiplication calculation expression and text representation contain "cut" because "cut" is associated with division calculation expression. Thus, the teacher has to spend three lessons for resolving this difficulty. The lessons have been performed from sixth lesson for level 4 continually and the subjects have worked on level 9 in eleventh lesson.

The result of average score and SD in usual problem solving, extraneous problem solving and problem-posing test are shown in Table 2. These scores are divided into experienced and inexperienced group of MONSAKUN that the subjects learn by problem-posing in the scope of one-step addition, subtraction or multiplication word problem. In addition to this result, the results of ANOVA in each test are shown in Table 3. There was an interaction in the score of usual problem-posing test between experience of MONSAKUN and pre-posttest ($p=.03$). So, we analyzed simple effect. There was a significant difference in the score of posttest between experienced group and inexperienced group ($F(1, 36)=3.193$, $p=.008$). This result suggested that it is effective for the subjects to experience the learning by using MONSAKUN for improving their usual problem solving performance. Next, there was a significant difference in the score of extraneous problem solving test between experienced group and inexperienced group ($p=.04$), and effect size is medium ($\eta^2=.10$). Also, there was a significant difference in the score between pretest and posttest ($p=.02$), and effect size is small ($\eta^2=.02$). In addition to this analysis in the score of extraneous problem solving test, we analyzed the correlation between the pretest score and the difference posttest and posttest score. In this result, there are a negative correlation between them (Spearman's rank-correlation coefficient, $|r_s|=59$, $p=2.5E-06$). These results suggested that the lesson by using our learning environment promote the subjects to improve their problem structure, in particular, more effective to the subjects who the score of extraneous problem solving test is lower. MONSAKUN is more effective for the subjects who have experienced MONSAKUN to comprehend the problem structure particularly. This result same as the result of experimental use by MONSAKUN Touch 2 (Yamamoto, et al, 2013) so this effect is that the main effect of learning by problem-posing on MONSAKUN Touch. Last, there was no significant difference in the number of posed problem between experienced and inexperienced group. But, there was a significant difference between pretest and posttest ($p=.005$), and effect size is medium ($\eta^2=.07$). These results suggested that MONSAKUN is useful for the subjects to improve their problem-posing performance regardless of whether the subjects have experienced MONSAKUN.

Table 2: Result of Each Pretest and Posttest in experienced group (N=18) and inexperienced (N=20).

<table>
<thead>
<tr>
<th>Test</th>
<th>Experience of MONSAKUN</th>
<th>Pretest</th>
<th>Posttest</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>M</td>
<td>SD</td>
</tr>
<tr>
<td>Problem-posing</td>
<td>experienced</td>
<td>2.50</td>
<td>1.57</td>
</tr>
<tr>
<td></td>
<td>inexperienced</td>
<td>2.10</td>
<td>1.45</td>
</tr>
<tr>
<td>Usual Problem solving</td>
<td>experienced</td>
<td>13.61</td>
<td>1.34</td>
</tr>
<tr>
<td></td>
<td>inexperienced</td>
<td>13.30</td>
<td>1.52</td>
</tr>
<tr>
<td>Extraneous Problem Solving</td>
<td>experienced</td>
<td>10.83</td>
<td>2.14</td>
</tr>
<tr>
<td></td>
<td>inexperienced</td>
<td>8.80</td>
<td>3.54</td>
</tr>
</tbody>
</table>
Table 3: Two factor ANOVA of Each Pretest and Posttest.

(a) Result of the score of usual problem solving test

<table>
<thead>
<tr>
<th>factor</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>experienced × inexperienced group</td>
<td>16.02</td>
<td>1</td>
<td>16.01</td>
<td>4.56</td>
</tr>
<tr>
<td>pre × post-test</td>
<td>0.06</td>
<td>1</td>
<td>0.06</td>
<td>0.05</td>
</tr>
<tr>
<td>interaction</td>
<td>7.01</td>
<td>1</td>
<td>7.01</td>
<td>5.32</td>
</tr>
<tr>
<td>total variation</td>
<td>196.88</td>
<td>75</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(b) Result of the score of extraneous problem solving test

<table>
<thead>
<tr>
<th>factor</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>experienced × inexperienced group</td>
<td>63.88</td>
<td>1</td>
<td>63.88</td>
<td>4.54</td>
</tr>
<tr>
<td>pre × post-test</td>
<td>10.74</td>
<td>1</td>
<td>10.74</td>
<td>5.63</td>
</tr>
<tr>
<td>interaction</td>
<td>0.74</td>
<td>1</td>
<td>0.74</td>
<td>0.39</td>
</tr>
<tr>
<td>total variation</td>
<td>651.41</td>
<td>75</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(c) Result of the score of problem-posing test

<table>
<thead>
<tr>
<th>factor</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>experienced × inexperienced group</td>
<td>7.47</td>
<td>1</td>
<td>7.47</td>
<td>2.28</td>
</tr>
<tr>
<td>pre × post-test</td>
<td>12.98</td>
<td>1</td>
<td>12.98</td>
<td>9.19</td>
</tr>
<tr>
<td>interaction</td>
<td>0.98</td>
<td>1</td>
<td>0.98</td>
<td>0.70</td>
</tr>
<tr>
<td>total variation</td>
<td>190.04</td>
<td>75</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* p<.05, ** p<.01

5. Conclusion

In this paper, we have described the model of problem and problem-posing in one-step multiplication or division arithmetic word problem, the development of the interactive environment for problem-posing based on the model, and the results of its practical use. We analyze the problem structure as processing information and realize the development of interactive environment for learning by problem-posing based on its problem structure continually. This learning environment called MONSAKUN Touch. Until now, MONSAKUN Touch is introduced by elementary school teacher into first and second grade arithmetic class on an elementary school for practical use. MONSAKUN Touch is developed for learning by problem-posing in the scope of one-step addition, subtraction or multiplication arithmetic word problem. Therefore, as the next step, we designed and developed the learning environment by posing problem that can be solved by one-step multiplication or division word problem. In order to realize this system, firstly, we have mentioned that three quantities and its relation define one-step multiplication or division word problem. These three quantities are called base quantity, proportion and compared quantity. Its relations are "Base quantity multiplied by proportion is compared quantity". At the second step, the problem-posing based on this problem structure and the diagnosis and feedback of posed problem are defined. The levels of assignment are designed by this problem structure as the learner can acquire the problem structure gradually. After that, we have developed learning environment for problem-osing as sentence integration. This environment consists of MONSAKUN Touch 3 and MONSAKUN Analyzer 3. The learner can exercise the problem-posing on MONSAKUN Touch 3 and MONSAKUN Analyzer 3 provides the visualized student's learning data on MONSAKUN Touch 3 to the teacher. Lastly, an eleven lesson experimental use is reported. The results of brief analysis suggested that the third grade students who have learned by using MONSAKUN in the past time are improved their problem solving performance and sophisticated their acquired problem structure. In addition to this result, the third grade students who didn't acquire the problem structure well are improved their problem-posing
and sophisticated their acquired problem structure. This result same as the result of experimental use by MONSAKUN Touch 2 so it is suggested that this result is main effect of learning by problem-posing on MONSAKUN Touch.

As our future works, we need to verify the quantity of the effect to high group by using MONSAKUN Touch 3. Furthermore, we should perform the practical use for problem-posing in one-step addition, subtraction, multiplication or division to fourth grade students of an elementary school continuously.

References

Revealing Students' Thinking Process in Problem-Posing Exercises: Analysis of First Sentence Selection

Nur HASANAH*, Yusuke HAYASHIa & Tsukasa HIRASHIMAa
aDept. of Information Engineering, Hiroshima University, Japan
* nur@lel.hiroshima-u.ac.jp

Abstract: We have developed a computer-based learning environment called MONSAKUN to realize learning by problem-posing, where students select and arrange several sentence cards to pose arithmetical word problems. We call this type of problem-posing “sentence integration”. As the next step of MONSAKUN development, we have been analyzing the sentence selection process which is considered to reflect students' thinking process. In the first step of analysis, we focused on the first sentence card selected in the process of posing a problem. We found that the selection changed based on different type of approach, type of story and students’ exercise experience. This result is an important step towards building elaborate process model of problem-posing and adaptive support of the process.

Keywords: Arithmetical word problems, learning analytics, problem posing, reverse thinking problem, sentence integration

1. Introduction

1.1 Background

Two activities that have been identified to be central themes in mathematics education are problem posing and problem solving. Problem solving practice, as the most popular way of teaching the solution method, has been long integrated into school mathematics (Stanic & Kilpatrick, 1988). Problem posing practice involves the generation of new problems in addition to solving pre-formulated problems (English, 1997; Silver & Cai, 1996). Although learning by problem posing has been suggested as an important way to promote learner’s understanding (Ellerton, 1986; Polya, 1957), it was not until recently that the recommendations for the reform in mathematics education suggested the problem posing inclusion in students' activities (NCTM, 2000). Several investigations of various aspects of problem posing activities have been conducted as more educators and researchers realized its importance in mathematics education (English, 1998; English, 2003).

One of the most important issues in learning by problem posing is the way to assess and give feedback to posed problems. In traditional problem posing method, teachers and students were faced with difficulties to conduct the learning activities effectively. It is not easy for students to pose mathematically correct problems in a given time, especially students in lower grade of elementary school. Teachers were having problems to assess and give feedback to the wide variation of problems that students pose in a given time of class activity. The inefficiency of time and available method made problem posing activity less attractive for most mathematics educators.

In order to realize learning by problem-posing in a practical way, we have been investigating a computer-based learning environment to assess and give feedback to problems posed by students (Nakano, Hirashima & Takeuchi, 1999; Hirashima, Nakano & Takeuchi, 2000; Nakano, Hirashima & Takeuchi, 2002; Hirashima et al, 2007; Kurayama & Hirashima, 2010). The software, named MONSAKUN (means “Problem-posing Boy” in Japanese), provides an interactive support for learning arithmetical word problems solved by one operation of addition/subtraction.

The interface of an assignment in MONSAKUN is explained in Figure 1. A learner is provided with a set of sentence cards and a numerical expression, and then he/she is required to pose an
arithmetical word problem using the numerical expression by selecting and arranging appropriate cards. Although learners do not create their own problem statements, they are required to interpret the provided sentences and integrate them into one problem, which is essentially the same as ordinary problem-posing activity. Hirashima & Kurayama (2011) call this style as “problem-posing as sentence-integration” and assert that this integration process is an essential activity in learning.

Hirashima & Kurayama (2011) call this style as “problem-posing as sentence-integration” and assert that this integration process is an essential activity in learning.

Figure 1. Interface of MONSAKUN

The practical use of MONSAKUN at several elementary schools has been reported in previous studies (Hirashima et al., 2008; Kurayama & Hirashima, 2010; Hirashima & Kurayama, 2011; Yamamoto et al., 2013). Through the analysis of pre-test and post-test of high-score group and low-score group of the students, effect of learning by problem-posing with MONSAKUN was investigated. It has been confirmed that the problem-posing exercise is effective to improve both problem-posing and problem-categorization abilities. Furthermore, after long term use of MONSAKUN in an elementary school, the result showed that both the students and teachers enjoyed using this system continuously and considered it useful for learning.

1.2 Purpose

One important direction in investigation of problem posing activities is to examine thinking processes related to problem posing (Brown & Walter, 1990). As the next step of MONSAKUN development, the purpose of this study is to examine learners’ problem-posing process and to develop technologies for identifying learners’ thinking process. By identifying learners’ thinking process, we will be able to provide a better individualized feedback based on understanding of each learner.

Through previous practical use, we observed different ways of sentence selection in problem-posing process by the students. We assume that it is caused by the different way of thinking depending on the nature of problems and learner's understanding. Therefore, by examining the selection process of sentences, we aim to infer about a learner’s thinking process in problem-posing.

While it is difficult to trace thinking process in a free problem posing activity, we can trace learners’ card selection in MONSAKUN which can be considered to reflect their thinking process. Problem posing in MONSAKUN is defined as integration of provided sentences into one problem. Learner’s assignment is to choose appropriate cards from several sentence cards provided by the system in order to fill the requirement of numerical expression and story type. This can be considered as search problem. Figure 2 illustrate a search space of an assignment in MONSAKUN which provides six sentence cards. The search space is a tree structure of combination of cards. Here, the root is the starting point and the numbers represent ID of cards, for example, the starting point is empty and the combination of cards 1, 2 and 3 indicates the correct answer. The nodes and arrows with bold line represents the paths of the learner’s card selections during his problem posing activity.
committed mistakes twice and then got the correct answer. As shown in this figure, problem-posing in MONSAKUN is defined as a search problem in the structure of transition of card combinations. The rules for valid combination of the sentence cards are explained in the next section.

Figure 2. Example of learner’s card selections shown in a graph

In this study, we examine how learners pose arithmetical word problems as sentence integration on MONSAKUN. Our assumption is learners do not choose sentence cards randomly - they arrange sentence cards based on some sort of thinking. In the analysis, as the first step toward analyzing problem-posing activity, we especially focus on what kind of sentence card was firstly selected by the learners.

The composition of this paper is as follows. The next section gives an overview of MONSAKUN and the definition of two types of problems: forward-thinking and reverse-thinking problem. Section 3 describes data about initial card selection by learners and discusses what happens in problem posing on MONSAKUN. Finally section 4 concludes this paper and shows some prospects for future study.

2. Problem Posing Activity in MONSAKUN

2.1 Categorization of Problem-posing Exercises

In arithmetical word problems, sentences are divided into two types: existence sentence and relational sentence. An existence sentence represents a number of single objects. A relational sentence contains keyword that represents a story type. An arithmetic word problem of binary operation is integration of two existence sentences and one relational sentence.
There are four types of story in arithmetic word problems of addition and subtraction: 1) combination, 2) increase, 3) decrease, and 4) comparison (Riley, Greeno and Heller, 1983). In MONSAKUN, the differences among them are defined as differences of integration of sentences. For example, a decrease story type problem is composed as follows:
   a) There are seven apples (existence sentence),
   b) Several apples were eaten (relational sentence that contains decrease story type), and
   c) There are three apples (existence sentence).

2.2 Forward-thinking and Reverse-thinking Problem

An arithmetical word problem includes two kinds of numerical relations: story operation structure and calculation operation structure. Story operation structure is the equation expressing the numerical relation according to the story, while calculation operation structure is the equation used to derive the required number in the assignment.

Based on this relation, there are two groups of problem in arithmetical word problems: forward-thinking problem and reverse-thinking problem. In forward-thinking problem, a story represented in the problem has the same structure with the calculation to derive the answer, while in reverse-thinking problem, the story and the calculation operation structures are different (Hirashima and Kurayama, 2011).

For example, in the following problem:

There are seven apples. Three apples were eaten. There are several apples.

How many apples are there?

Based on the sentence “Three apples were eaten”, we understand that the story focuses on “decrease” number of an object. The story operation structure is “7-3=?”, and the calculation structure is also “7-3(=?)”, which can be found easily by reading the story in order from the first sentence. Since the two structures are the same, this type of problem can usually be solved easily by the learners.

Meanwhile, in the following problem:

There are seven apples. Several apples were eaten. There are three apples.

How many apples were eaten?

We can derive the story operation structure as “7 - _ = 3”, and the calculation operation as “7 - 3 = _”. Since the two structures are different, a learner is required not only to understand the story but also to derive the calculation operation structure from the story. This type of problem is called “reverse-thinking problem”.

2.3 Task Model of Problem-Posing

Based on the consideration of problem types, we have proposed a task model of problem posing as sentence-integration shown in Figure 3 (Kurayama & Hirashima, 2010). There are four main tasks in problem posing activity: (1) deciding calculation operation structure, (2) deciding story operation structure, (3) deciding story structure, and (4) deciding problem sentences. A learner should complete these tasks to pose a correct problem, although the execution procedure of the tasks is not decided in the model.

In the first step of MONSAKUN, subtraction or addition is selected as a calculation operation structure. In the second step, a story operation structure is decided. For example, for subtraction, four story operation structures can be selected. Among them, only one story operation structure is the same with the calculation operation structure (subtraction), and two of them have completely different story operation, that is, addition. Because this is an abstract transformation, it is often very difficult for learners to decide.

The next task of deciding story structure involves selection from four types of story: combination, increase, decrease, or comparison problem. Each type of story has its own structure, as explained in the sections above.

In the last task of deciding problem sentences, sentences are put into the story structure following the story operation structure. This task is divided into three more tasks: deciding sentence structure, deciding concept structure and deciding number structure. Deciding sentence structure is to select and order sentences following the story structure. For example, if the story structure is decrease
problem, learner should make a sentence structure composed of an existence sentence, a decrease type of relational sentence, then followed by another existence sentence.

Figure 3. Task model of problem-posing as sentence integration

For the decision of concept structure, concepts dealt with the problem are decided. For example, if the problem requested learner to answer about the total number of apples and oranges, then the sentences should be dealt with apples and oranges as the concept. For the decision of number structure, the numbers dealt with the problem is decided. In arithmetic word problems, a negative number should not be used.

3. Analysis of MONSAKUN Log Data

In this section, the analysis of MONSAKUN log data from an experiment of MONSAKUN used by eleven undergraduate students from Faculty of Education is reported. Although MONSAKUN is intended for elementary school students, the subjects of this experiment are undergraduate students. The reason is that undergraduate students are supposed to be able to solve both forward-thinking and reverse-thinking problems rather easily, because they have already understood the structure of simple arithmetic word problems. They are only expected to learn how to make problems through the use of MONSAKUN.

On the other hand, elementary school students firstly learn about the problem structure through the use of MONSAKUN before they become able to pose problems, which takes several times of class schedule. Because the undergraduate students do not need to learn but only to recognize the problem structure, they are expected to show clearer changes in thinking process towards different problem types than elementary school students. For this reason, our study analyzed data from the experimental use of MONSAKUN by university students as the subjects. In the experiment, the subjects are firstly given explanation about the software, and then posed problems in a given time.

Our aim in this study is to examine learners’ way of thinking from selection of sentence, especially the first selected sentence in each assignment. We analyzed the subjects’ log data in assignments at Level 1 and Level 5 which require the subjects to pose forward-thinking problems and reverse-thinking problems, respectively. Both levels consist of 12 assignments that include four types of stories: combination, increase, decrease, and comparison. Each type of story has three assignments. Subjects carried out the assignments in order, and they can only move on to the next assignment when the current assignment has been answered correctly.
3.1 Difference in First Selected Card between Level 1 and Level 5

Figure 4 shows log data in Level 5 Assignment 1 mapped on a graph. Black nodes and links represent the ones selected by the subjects, while gray ones represent the ones not selected. In this experiment, not all paths were observed in subjects’ selection. Subjects only took some particular paths, which show that the card selections are not random. In addition, focusing on the card firstly selected by subjects, most of them chose Card 4 (8 subjects out of 11). In this assignment, there is a decided tendency based on some sort of thinking. If we can clarify the tendency for learners to choose a specific first sentence card, it will be useful to diagnose learners’ understanding. Therefore, as the initial step of analysis of students' thinking process, this study aims at revealing the characteristics of first selected sentence card.

The sentence cards in MONSAKUN contain different number according to the numerical expression in the given assignment. For example, in an assignment “Make a story problem about ‘how many are the difference’ that can be solved by 7 – 3.”, the required calculation expression is “7 – 3 = _”. In MONSAKUN, several sentence cards with numbers are provided to the users. The cards are distinguished by the order of numbers in the required calculation expression. If a card contains the first number in the required expression, for example, 7 in the example above, it is called “first number card”. Similarly, if it contains the second number or the third number, it is called “second number card” or “third number card”, respectively. One of the numbers in every assignment is an unknown number, which is represented by the blank mark.

Figure 4. University students’ card selection in Level 5 Assignment 1 shown in graph
From the analysis of all subjects’ first card selection in Level 1 and Level 5 assignments, we found that the proportion of each sentence card to be selected firstly is entirely not even. Table 1 shows the proportion of first card selected by subjects in Level 1 and 5. In every assignment, only one or two cards are significantly chosen by them. We found that there is a bias against first selected card. This finding proves our assumption that subjects did not choose a card randomly, but with some sort of approach.

Table 1 Percentage of first selected card by the subjects

<table>
<thead>
<tr>
<th>Type of first selected card</th>
<th>Level 1 (%)</th>
<th>Level 5 (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>First number card</td>
<td>91.8</td>
<td>58.7</td>
</tr>
<tr>
<td>Second number card</td>
<td>3.3</td>
<td>16.5</td>
</tr>
<tr>
<td>Third number card (blank mark)</td>
<td>4.9</td>
<td>24.8</td>
</tr>
</tbody>
</table>

Furthermore, we found different trends of first card selection between Level 1 and 5. We presume that this difference appeared because subjects had different approach to pose either forward-thinking or reverse-thinking problems. In forward-thinking problem, the approach to order cards following the order of numbers in the numerical expression can be applied easily. However, in reverse-thinking problem they cannot pose problem with the same approach. This type of problem requires learners to think about the numerical relation in the given problem and reflect it to the choice of cards.

3.2 Change of Approach through the Exercise

In the previous section, we have presumed that subjects had different approach to solve forward-thinking and reverse-thinking problems. In this section, we would like to explain further how the subjects change their way of thinking during problem posing exercise by looking at the type of story, order of assignment, type of first selected card, as well as the type of sentence. We especially analyzed subjects’ selection in Level 5 assignments, where they posed challenging reverse-thinking problems.

Table 2 shows the characteristics of first selected sentence card from each assignment at Level 5 that has marginal or significant difference in number of selection from the average. These results were analyzed with binomial test to the amount of each card being firstly chosen or not in each assignment. Binomial test is an exact test of the statistical significance of deviations from a theoretically expected distribution of observations into two categories. Based on our assumption that students posed problems by selecting cards through a thinking process, we expect the distribution of firstly chosen cards to have a significant difference in comparison with other cards.

Table 2. Result of binomial test of first selected card in Level 5 assignments

<table>
<thead>
<tr>
<th>No</th>
<th>Type of story</th>
<th>Order of assignment</th>
<th>Type of first selected card</th>
<th>Type of sentence</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Combination</td>
<td>1&lt;sup&gt;st&lt;/sup&gt;</td>
<td>First number card</td>
<td>Existence</td>
<td>7.05*10&lt;sup&gt;-5&lt;/sup&gt; **</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>2&lt;sup&gt;nd&lt;/sup&gt;</td>
<td>First number card</td>
<td>Relational</td>
<td>1.88*10&lt;sup&gt;-7&lt;/sup&gt; **</td>
</tr>
<tr>
<td>3</td>
<td></td>
<td>3&lt;sup&gt;rd&lt;/sup&gt;</td>
<td>First number card</td>
<td>Relational</td>
<td>1.97*10&lt;sup&gt;-3&lt;/sup&gt; **</td>
</tr>
<tr>
<td>4</td>
<td>Increase</td>
<td>1&lt;sup&gt;st&lt;/sup&gt;</td>
<td>First number card</td>
<td>Existence</td>
<td>1.89*10&lt;sup&gt;-7&lt;/sup&gt; **</td>
</tr>
<tr>
<td>5</td>
<td></td>
<td>2&lt;sup&gt;nd&lt;/sup&gt;</td>
<td>Second number card</td>
<td>Existence</td>
<td>0.0504 +</td>
</tr>
<tr>
<td>6</td>
<td></td>
<td>3&lt;sup&gt;rd&lt;/sup&gt;</td>
<td>First number card</td>
<td>Existence</td>
<td>0.0504 +</td>
</tr>
<tr>
<td>7</td>
<td>Decrease</td>
<td>1&lt;sup&gt;st&lt;/sup&gt;</td>
<td>First number card</td>
<td>Existence</td>
<td>2.35*10&lt;sup&gt;-4&lt;/sup&gt; **</td>
</tr>
<tr>
<td>8</td>
<td></td>
<td>2&lt;sup&gt;nd&lt;/sup&gt;</td>
<td>Second number card</td>
<td>Existence</td>
<td>2.35*10&lt;sup&gt;-4&lt;/sup&gt; **</td>
</tr>
<tr>
<td>9</td>
<td></td>
<td>3&lt;sup&gt;rd&lt;/sup&gt;</td>
<td>Second number card</td>
<td>Existence</td>
<td>2.35*10&lt;sup&gt;-4&lt;/sup&gt; **</td>
</tr>
<tr>
<td>10</td>
<td>Comparison</td>
<td>1&lt;sup&gt;st&lt;/sup&gt;</td>
<td>-</td>
<td>-</td>
<td>0.0266 *</td>
</tr>
<tr>
<td>11</td>
<td></td>
<td>2&lt;sup&gt;nd&lt;/sup&gt;</td>
<td>Third number card</td>
<td>Relational</td>
<td>0.0266 *</td>
</tr>
<tr>
<td>12</td>
<td></td>
<td>3&lt;sup&gt;rd&lt;/sup&gt;</td>
<td>Third number card</td>
<td>Relational</td>
<td>0.0266 *</td>
</tr>
</tbody>
</table>

**: significant difference (p<.01), *: significant difference (p<.05), +: marginal difference (p<.1)
When firstly used MONSAKUN, subjects are given simple forward-thinking problems to pose at Level 1. From the analysis mentioned in Section 3.1, we found that they first simply chose a card with the first number in the required numerical expression (“first number card”), and then proceeded to choose other appropriate cards. This approach worked well for assignments in Level 1, where all of the assignments are forward-thinking problems.

When subjects arrived at the first assignment of Level 5, they initially approached the assignment with the same way of thinking in choosing the first sentence card. However, this did not work well, and they tend to make more mistakes than in the previous levels. We presumed that the subjects were aware that the previous approach did not work for reverse-thinking problems, because in the second assignment of Level 5 they tend to choose another type of card.

In a similar way, subjects changed their approach from the first assignment in a story type to the second and third assignment in the same story type. As shown in Table 2, in the first assignment in each type of problem, they generally took the simple approach to firstly select a “first number card” containing an existence sentence. Only in the case of comparison story there was no significant difference in cards selected by subjects in the first assignment. On the other hand, in the second and third assignments of the same type of story, they did not choose it as the first card. For example, in combination stories, most of them firstly did not select existence sentence, but “first number card” containing relational sentence. This is also the same as in the decrease story type.

Meanwhile, in the case of increase stories, we did not found any evident change between the assignments. At the second assignment they tend to select “second number card” containing existence sentence (shown by a marginal p-value), while at the third assignment they took the simple approach just like in previous levels of forward-thinking problems.

Furthermore, in comparison stories, there is no trend in first card selection at the first assignment. However, at the second and third assignment, there is a trend to select “third number card”, that is a blank mark, containing a relational sentence. Consequently, we observed that there is a change of approach in comparison story compared to the previous story types.

This leads to two findings about changes in subjects’ way of thinking through the exercises. The first one is that subjects change their approach to pose problems after they had experienced posing the same type of story. As shown in Table 2, trends of first card selection are different between the first assignment and the rest in the same story type. The next finding is that the change of approach depends on the type of story, as we can see that subjects made different first card selection in different story type.

Table 3. Average of steps and mistakes in Level 5 assignments

<table>
<thead>
<tr>
<th>Story Type</th>
<th>Average No. of Steps</th>
<th>Average No. of Mistakes</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Assignment</td>
<td>Assignment</td>
</tr>
<tr>
<td></td>
<td>1st</td>
<td>2nd &amp; 3rd</td>
</tr>
<tr>
<td>Combination</td>
<td>11.60</td>
<td>4.20</td>
</tr>
<tr>
<td>Increase</td>
<td>45.50</td>
<td>16.50</td>
</tr>
<tr>
<td>Decrease</td>
<td>24.90</td>
<td>16.30</td>
</tr>
<tr>
<td>Comparison</td>
<td>10.00</td>
<td>9.80</td>
</tr>
</tbody>
</table>

These changes of approach seem to bring a good effect to subjects’ thinking process in posing reverse-thinking problems. Our analysis of the average of steps and mistakes in Level 5 problems showed that in comparison to the first assignment of each story type, the average of steps and mistakes in the second and third assignments of the same story type are mostly decreased, as shown in Table 3.

4. Concluding Remarks

In this research, we have conducted analysis of MONSAKUN log data of university students’ problem posing activity to investigate their way of thinking in posing different types of arithmetical word problems. From the analysis, we found that the first sentence selected in each assignment were different in several ways. In forward-thinking problems, subjects generally used a simple approach to select
“first number card”. However, in reverse-thinking problems, they changed the approach to select “second number card” or “third number card”. Depending on the type of story and subjects’ exercise experience, they applied different approach of first card selection. These finding proves our assumption that learners who used MONSAKUN did not chose cards randomly, but with some sort of thinking process. Furthermore, we infer that learners who used MONSAKUN were able to recognize the differences in structure of problems depending on types of story, as they changed their approach to pose problems for different story types. The recognition of the difference is important for learners to understand the nature of arithmetic word problems.

For the next step of this research, we plan to perform the analysis to a larger data of MONSAKUN used by elementary school students to infer their thinking process of the same problem posing activity. The results will be used to make an elaborate process model of the problem-posing and adaptive support of the process.

References

Balance Control of Question-Posing Focusing on Learning Target Words on the Self-Study Material Contribution and Sharing System

Toshihiro HAYASHI\textsuperscript{a}, Yuji HIRAI\textsuperscript{b}, Kazuhiro URA\textsuperscript{a}, Akihiro IWAKI\textsuperscript{c}, Rihito YAEGASHI\textsuperscript{b}, Hiroshi MURAI\textsuperscript{c} & Hiroyuki TARUMI\textsuperscript{b}
\textsuperscript{a}Information Technology Center, Kagawa University, Japan
\textsuperscript{b}Faculty of Engineering, Kagawa University, Japan
\textsuperscript{c}University Consortium for E-Learning, Shikoku Center, Kagawa University, Japan
*hayashi@eng.kagawa-u.ac.jp

Abstract: Recently, education and learning by using e-Learning style have become popular. We have developed a self-Study Material Contribution and Sharing System called “S-Quiz” wherein students can pose questions and share them. Through the educational practice with S-Quiz, we found that balance of posed questions was not suitable to use for learning after question-posing. This paper describes balance control of question posing in order to solve this problem of unbalanced questions. This balance control aims to increase questions of which posed number is a few by system-centered navigation to students on S-Quiz. In addition, we propose a control method which is not biased to a part of students.

Keywords: e-Learning, Self-Study Material Contribution and Sharing System, S-Quiz, Balance Control of Question-Posing

1. Introduction

Recently, education and learning by e-Learning have become popular in especially higher education, job training, and so on. We have developed a self-Study Material Contribution and Sharing System called “S-Quiz” wherein students can pose questions and share them(Mizuno et al, 2007, Hayashi et al, 2008). S-Quiz is categorized into educational system by using question-posing. There have been developed various kinds of such question-posing systems(Nakano et al 2002, Yu et al 2005, Takagi and Teshigawara 2006, Hirai and Hazeyama 2007, Denny et al 2008).

We used S-Quiz in home work for a special subject of university. In the home work, each student could freely pose questions about several important keywords which s/he understands well. We call the important keywords which can be selected in a lecture LTWs: Learning Target Words in this paper. In this educational practice, we found students posed many questions for several LTWs as a positive result but other several LTWs have no posed question as a negative result. We can think that the unbalance of question-posing can happen by allowing such style of question-posing mentioned-above.

Posed questions are expected to be shared and be used their learning after question-posing. Therefore, unbalanced question sets are not suitable for student to use for their learning. In order to avoid the unbalance of questions, we try to add a balance control function into S-Quiz. In this balance control function, “good balance of question posing” is defined that the number of questions about each LTW is bigger than the threshold value (for example, minimum number of posed-questions). Based on this idea, we set following conditions:

(1) Balance Control function navigates student to pose question about LTWs of which the number of posed questions is smaller than the threshold value, and
(2) The navigation is not concentrated to a part of students.

In this paper, we propose two balance control methods: basic method and improved method. Especially, explain about the simulation result of the improved method and show this method satisfy the conditions (1) and (2) on simulation level.
2. S-Quiz

S-Quiz provides a learning environment wherein students can pose questions freely and share the question with other students. Fig.1 shows a snapshot of question posing interface of S-Quiz. After choosing question category, students can make a question by inputting a question text, one correct answer, and three incorrect answers on the question posing interface and then post it. In addition, students can input hint information of questions by text and image data. In the case of Fig.1, the student poses a question about ENIAC. Generally, ENIAC becomes a LTW which can be selected in the genre about computer system. As for the genre, student can select a proper genre from the pull down menu. The set of genre is chosen by the teacher in advance.

![Figure 1: Question posing interface of S-Quiz](image1.png)

Fig.1 Question posing interface of S-Quiz

Fig.2 shows a snapshot of question answering interface of S-Quiz. Students can also share all questions posed by other students and use them for students’ free learning. Beside the multiple choice question, the genre, the maker (student) and the evaluations of the questions quality are displayed. As for the evaluation, students can vote three kinds of evaluations: “Good”, “Average” and “Not good” to questions which the student has answered. Hint message and image can be also displayed if they are prepared. Basically, the student can answer the question by selecting a correct answer candidate from four choices, but s/he can select “Pass” to obviously inform that s/he does not know.

![Figure 2: Question answering interface of S-Quiz](image2.png)

Figure 2: Question answering interface of S-Quiz
3. Balance control of question-posing

3.1 The reason of unbalance of question posing

In S-Quiz, student’s actions are mainly classified into question-posing and question-answering. LTWs are important words in a lecture. Therefore, it is ideal style of question-posing in S-Quiz that all students make questions about all LTWs. However, we should consider the other factors: the number of students, understanding levels for each LTW, and so on. As for the number of students, if it is large number, the number of posed questions also becomes large. Sometimes it becomes exceed to use. In addition, the possibility of low quality of questions becomes high, because it is not guaranteed that all students can good quality of question for every LTW from the viewpoint of understanding levels for each LTW. Therefore, it is a realistic and loose solution for answering to the mentioned-above that each student poses questions about several LTWs which s/he understands well. In contrast, the unbalance of question-posing can happen by allowing such style of question-posing.

3.2 Conditions of Balance control

Balance control of question-posing is to increase question-posing about LTWs of which the posed number is smaller than the threshold value in order to avoid unbalance of question-posing. S-Quiz focuses on such LTW and navigate student to pose question about the LTW. This control makes question-posing balance good with satisfying the following conditions: (1) Balance Control function navigates student to pose question about LTW of which the number of posed questions is smaller than the threshold value, and (2) The navigation is not concentrated to a part of students.

3.3 Basic method

Considering only balance of question-posing, we can just focus on the condition (1). As for LTWs of which the number of posed questions is smaller than the threshold value, S-Quiz navigates students to pose questions about the LTW. We call this navigation as “basic method.” The followings are concreate process of basic method:

(BM1) Estimation of the number of questions which students will pose from now,
(BM2) Calculation of probability that each LTW is included in questions, and
(BM3) Decision of LTWs that navigation of question-posing is needed by using results of the above mentioned (BM1) and (BM2).

The basic method is simple to be implemented. However, navigation based on the basic method tends to become frequent at the late stage of question-posing. In other words, the navigation is concentrated to a part of students who pose questions at such stage. Unless students accept the navigation, the balance of question-posing cannot be improved.

3.4 Improved method

Focusing on both conditions (1) and (2), we propose “improved method” for balance control of question-posing. This improved method positively navigates students even at early stage of question-posing. In order to realize this feature, the necessary number of posed question of each LTW is calculated automatically according to the situation. If the current number of posed question about a LTW is smaller than the necessary number, navigation works for question posing about the LTW. To fulfil the condition (2), the improved method activates navigation at early stage of question-posing even when navigation is not needed.

3.5 Discussion

Generally, it can be useful for keeping the balance of numbers to use both upper and lower limits. However, both basic method and improved method use only a lower limit. If upper limit is set, it can occur that a student is not allowed to pose the question about LTWs on upper limit. This is a bad situation for motivation of the student and also reduces the possibility that the student can pose
questions. The root purpose of balance control is to guarantee the minimum number of posed questions about each LTW. If the root purpose is satisfied, there is no need to set the upper limit. In addition, lower limits control by both methods loosely restricts exceed question-posing for each LTW. Through these considerations, the basic and improved methods put focus on only lower limits.

4. Simulation experiment

We did simulation experiments for evaluating performance of the improved method. In this section, we describe the outline of experiment, the simulation results, and discussion about the results.

4.1 Outline of simulation

Based on various kinds of data about educational practice with S-quiz, we assume student behavior and tendency of question-posing as followings:
(a) Students pose the required number (for home work) questions at once,
(b) Students do not use LTW which they have used before, and
(c) Students pose a question by using a LTW.
In addition, the following feature is included for simplification of group level behavior of students:
(d) Question-posing tends to be distributed and to be not overlapped.
Fig. 3 shows illustrated figure about the features (a) and (d).

![Figure 3 Question-posing tendencies of all students](image)

4.2 Simulation results

In this simulation, we set parameters as Table 1. The threshold value means the minimum number for each LTW. We obtained the results from the large number of simulation with the parameters and calculated the average.

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Simulation value</th>
</tr>
</thead>
<tbody>
<tr>
<td>the number of students</td>
<td>80</td>
</tr>
<tr>
<td>the number of question posing per one student</td>
<td>10</td>
</tr>
<tr>
<td>the number of LTW</td>
<td>20</td>
</tr>
<tr>
<td>the threshold value</td>
<td>20</td>
</tr>
<tr>
<td>the probability of acceptance ratio to navigation</td>
<td>50%</td>
</tr>
</tbody>
</table>
Fig. 4 shows the number of posed questions about each LTW. The number of each LTW shows tendency of use. Small number means question of the LTW is easy to be posed. In contrast, big number means question of the LTW is hard to be posed. Therefore, it becomes high possibility that the number of posed question about LTW with big number is not over the threshold value without the navigation.

![Figure 4 The number of posed questions about each LTW (by Improved method)](image)

In addition, Fig. 5 shows the number of navigation to each student. The number of student ID shows order of question-posing. For example, student 2 poses questions after student 1 poses questions (refer to Fig. 3). This means that students with big number pose questions at late stage. Such students have to pose questions under frequent navigation.

![Figure 5 The number of navigation to each student (by Improved method)](image)

4.3 Discussion

As for simulation results, we can find the every number of posed-questions for each LTW is over the threshold value (the minimum number) from Fig. 3. In addition, we can find the number of navigation is under two times and there is no big difference of navigation among all students from Fig. 4. We can say the improved method satisfies the condition (1) and (2). We also did simulation experiments with various parameter settings and obtained good results for the improved method. However, we omit those results because of paper length limitation.

5. Summary

In this paper, we explained about development and the educational practice issue of S-Quiz: a self-Study Material Contribution and Sharing System wherein students can pose questions and share them. In S-Quiz, each student is allowed to pose questions about several LTWs which s/he understands.
well. S/he does not have to pose questions about every LTW. In contrast, the unbalance of question-posing can happen by allowing such style of question-posing. Then, we proposed two methods for balance control of question-posing focusing on LTWs and simulation experiments. From the experiment we found our proposed method fulfil the following conditions: (1) Balance Control function navigates student to pose question about LTW of which the number of posed questions is smaller than the threshold value, and (2) The navigation is not concentrated to a part of students. The performance evaluation at real educational practice is remained.

Acknowledgements

This work was supported by MEXT (The Ministry of Education, Culture, Sports, Science and Technology) KAKENHI 26330401.

References

Development of a Customized English Learning System based on Augmented Reality Technology

Gwo-Haur Hwang1*, Beyin Chen2 & Hen-Lin Huang2
1Department of Information Networking and System Administration, Ling Tung University, Taiwan
2Department of Information Management, Ling Tung University, Taiwan
*ghhwang@teomainl.ltu.edu.tw

Abstract: The augmented reality (AR) is a kind of technology that can combine the virtual information and the real image. Many scholars indicate that AR can effectively enhance learners' learning motivation and effectiveness. This study proposes an English learning system based on AR technology. For customized design, the system adds the functions of related words and 3 learning scopes: phrases, and sentences, related words. In addition, an easy mode and an advanced mode are provided. These functions are less considered in the past and will also be regarded as the basis of future personalized preference analysis. Through the system, it is hoped that the learning motivation and learning effectiveness of learners can be enhanced. In the future, personalized preference will be analyzed according the collected historical data of learners’ behavior.

Keywords: customized design, English learning, augmented reality, personalized preference

1. Introduction

With the advances of sciences and technologies and the rise of the concept of the global village, people around the world can cross national boundaries through networks and communicate in the aspects of information and culture. Among these developments, the language is essential as a communication bridge. In the global trend of internationalization, learning English is no longer just the needs of individuals only, but also a trend of into a new era (Hwang, Lee, Hwang, Huang, Lin & Cai, 2013). Taiwan is an EFL (English as Foreign Language) English learning environment. In Taiwan, a formal exposure to English is until entering elementary schools (Lu, 2012). In the traditional teaching environments, through explaining the rigid text of the textbooks by teachers, the English learning can be conducted, which is passive learning (Savigan, 1988). In other words, the teachers just unilaterally give students the knowledge. In addition, due to the restrictions of the teaching time, students usually recite the knowledge, so that the knowledge cannot be applied to everyday life (Brown, Collins & Duguid, 1989).

Looi, Seow, Zhang, So, Chen and Wong (2010) have used a context-aware ubiquitous learning system to record the situation of learners. It allowed teachers to rapidly understand students' learning process and individual differences and to develop better teaching methods. By this, students can get the most necessary information in a timely and appropriate manner (Chen, Lien & Lu, 2009). The development process of context-aware technology first is RFID (Radio Frequency Identification) technology, which can complete a non-contact identification process by using the RFID reader to induce the RFID tags (Landt, 2005). Although the reliability and the identification speed of RFID technology are very high, but most smart phones and tablets do not support the reading of the RFID tags. In addition, the cost is also much more expensive than QR (Quick Response) codes. Thus, in the field of context-aware ubiquitous learning, QR Code technology has gradually replaced RFID technology. However, the virtual information and reality information provided by these two technologies are completely separated. This will result in that the users do not have coherent information when they receive information. According the "Spatial Contiguity Principle" and "Temporal Contiguity Principle" of 12 multimedia design principles proposed by Mayer (2009), the learning effectiveness of learners can be increased if at the time of scanning real objects, the
corresponding and associated information can be immediately generated beside the objects.

The augmented reality (AR) is a kind of technology that can combine the virtual information and the real image (Azuma, 1997). The technology allows that the learning process can more meet the above principles proposed by Mayer. Billinghurst (2003) has pointed out that AR can provide unique educational benefits. First, the use of AR as teaching aids allows learners smoothly to interact with virtual objects (interactive learning concept) in virtual and real environments. Second, the use of AR as teaching aids will extend as a new teaching and learning strategies. The learning mode can be conducted even if the students do not have any computer experience. Finally, AR has the characteristics that let learners be immersed in learning contents. It allows that learning is no longer just to face boring textbooks. Lai, Hwang, and Chen (2012) have pointed out that AR can indeed effectively enhance students’ learning motivation. Therefore, if a language learning system is built through AR technology, learners will be able to scan the learning objects directly in the real scene. The real-time multimedia information or teaching materials can be obtained, without destroying the original scene, but also saving the additional cost of teaching objects.

In addition, past researcheres have indicated that learners like to have options to set the functions according to their own preference and actual situation especially when the learning environment has many variables (Mitchell, Chen and Macredie, 2005). Therefore, this study proposes a customized AR English learning system. It is different from the general paper textbooks or traditional media textbooks which are just flat texts or pictures and lack context and interactivity. We expect that the customized AR English learning system can increase the learning wishes of learners through the good interactivity of AR technology. In addition, it can record the learning process which provides the future teaching reference for teachers.

2. Literature Review

2.1 The problems of English learning

In the past, second/foreign language learning relied upon teachers’ lecturing to explain the learning materials of textbooks (Savignon, 1988). In other words, because the teaching time is limited, traditional teaching is considered unable to enhance learning motivation and interest. It may result in the students’ to learn English by a rote style rather than by a style of increasing knowledge and problem-solving ability (Brown, Collins, & Duguid, 1989). Looi, Seow, Zhang, So, Chen and Wong (2010) have indicated that if the technologies of context-aware and ubiquitous learning are imported in teaching, learning will be able to be conducted at any time and any place. Thus, the role of teachers will transfer from imparter to guider. Teachers will guide students to learn actively and attract their attention so that the learners' ability of observation of the real world and the ability to actually solve problems can be enhanced. (Chen, Lien & Lu, 2009). Among the technologies of context-awareness, RFID and QR code technologies have the disadvantages of information discontinuous problem. Therefore, this paper uses AR technology to implement our learning system.

2.2 The applications of AR technology on education

AR is to import the images, objects, and scenes generated by computers to the real environment. Its purpose is to enhance the effect of perception. That is, the virtual objects are added to the real environment. This technology must have 3 characteristics: "combine the virtual and the real world", "be able to interact immediately", "be necessary in 3D space" (Azuma, 1997). Milgram, Takemura, Utsumi and Kishino (1994) regard real and virtual environments as a closed set as shown in Figure 1. The left is a purely real environment and the right is a purely virtual environment. The virtual reality attempts to replace the real world, while the augmented reality is to augment the virtual picture generated by computers into the real environment.
AR is currently widely used, for example, education, medical science, military training, engineering, industrial design, art, entertainment, etc. (Azuma, 1997). Among these applications, the application of AR on education has obtained the attention of many scholars. The reason is that the traditional way of multimedia learning lacks immediacy and interactivity. During the learning, the virtual and real information is separated. This disadvantage has been greatly improved after the rise of AR technology. Therefore, there are many scholars applying AR technologies to mobile navigation and context-aware ubiquitous learning.

There are many applications of AR technologies on mobile navigation. For example, Kuo (2008) applied AR to "epidemic battle Camp" exhibition at the National Science and Technology Museum. It allowed the audience to interact with objects in a 3D environment. The objects may be not able to actually take to the exhibition; may be not able to let the audience touch; or may be necessary to be amplified. Thus, the better communication effectiveness can be achieved. In addition, the stay time of visitors in front of display units can be extended. Sejin and Woontack (2009) proposed the guide learning of context-aware applied in the Museum of Art. It allowed users to follow the guiding indicator to deepen the art articles. In addition, according to different users, it provided personalized guidance interface. However, this study did not carry out effective analysis. Lin, Tang and Peng (2011) also used AR technologies to build digital teaching materials of arts and humanities learning of an elementary school. The results showed that the learners felt comfortable and easy when learning and operating AR aids. In addition, they have also maintained a certain degree of concentration and a sense of curiosity to AR aids. Wernhuar and Ou (2012) proposed an AR system of butterfly virtual ecological learning environment used on smart phones. The system was easy to maintain and solved the problem of insufficient butterfly species. The experimental results showed that the use of AR technology can improve the learning effectiveness. Chen and Tsai (2013) also applied AR to a library of an elementary school. They explored the effect of gender, prior knowledge and cognitive style on learning. The results found that AR learning way can enhance the overall learning satisfaction of learners. In summary, AR applied to mobile navigation can increase the learning willing and motivation of learners, but also enhance the learning effectiveness.

Many studies have also applied the AR technology to classroom teaching and reading. For example, Dünser and Hornecker (2007) studied and looked at the children' learning status of reading textbooks. They explored how for children of 6 to 7 years old to operate interactive teaching media. The results showed that a rich interface may increase the willingness of students to learn. Amir and Vinceet (2012) proposed a system to assist construction engineering students with AR technology to simulate the actual construction. Thus, the shortcomings of traditional construction which cannot practice actually can be improved. Under the simulation of AR environments, the students can conduct learning according to the real situation. Wu, Li, Yao and Pai (2012) designed an interactive AR system of chemistry experiment. Students can operate the picture cards of experiment equipment to complete a virtual chemistry experiment via AR technology. The difference between the traditional teaching style and the AR teaching style is that the students can watch the entire chemical reaction process through the 3D model. In addition, the experimental dangers that may occur can also be avoided. Chu and Lin (2013) have built an AR system which combined Kelly grid. It was used in natural science courses of an elementary school. The results showed that students' learning attitude was improved significantly. Chen, Zhao, Liu, Lin and Lu (2013) have built an AR system of gear teaching which imported the game concept. The experimental results indicated that the system can effectively enhance the concentration and the learning interest of students. In conclusion, we found that the use of AR technologies in the classroom can improve the shortcomings of the previous learning environment. In addition, it also enhanced the concentration of learners in classroom.

In addition, there are many applications of AR technology on language learning. For example,
Hsieh and Lin (2009) designed an AR system of English vocabulary learning which had immersive learning outcomes. The experimental results showed that the learners will be willing to use the system. Chang, Chen, Huang and Huang (2010) also built an AR game-based English vocabulary learning system. The above two English vocabulary learning systems are both based on AR technology, but the methods are different. The former is to use English vocabulary magic books so that learners can scan the learning objects in the books while the latter is to scan the 3D learning objects directly to conduct learning. The advantage of the latter is able to scan real objects in a real environment. It does not need extra teaching objects. So, this study adopts the latter method. Tsai, Li and Wu (2011) applied AR technology to Chinese learning. The learning system allowed foreigners to learn Chinese by combining learning objects of text boxes and the instant interactivity of AR technology.

However, the above teaching systems are all for vocabulary. The applications of related words or example sentences are lacking. In addition, it is inconvenient because additional objects are required. Therefore, in this study, we will develop a customized English learning system based on AR technology. It is hoped that through the combination of the real situation and the virtual information, the learning motivation and effectiveness of learners can be enhanced.

2.3 Customized design

Personalized design is to design system architecture or user interface according to the unique and special requirement of every user (Fink & Kobsa, 2000). Personalized design can be provided by 2 styles. One is customized and the other is adaptive. Regarding customized, users have the right of selection. That is, they are allowed to modify the content presentation styles, user interfaces and navigation tools themselves. Regarding adaptive, the system actively provides appropriate content presentation styles, user interfaces and navigation tools according to the observed user behavior. In other words, the former is user-oriented while the latter is system-oriented. Both have advantages respectively. The former gives users considerable options, but needs extra effort to choose their suitable manners. The latter although can decrease users’ loading, but may erroneously compute the users' preferences. However, the past researches have shown that most users prefer to have options (Mitchell, Chen and Macredie, 2005). Especially when the learning environment has many variables, the users like to set the content presentation styles, user interfaces, and navigation tools according to their own preference and actual situation.

3. System Implementation

3.1 The hardware architecture

Learners only need to install the customized AR English learning system to a smart phone or a tablet and then can scan the learning object using the smart phone or the tablet. When the identification is completed, the interactive learning in the real object and the virtual teaching material will be conducted. The operation of the system is very easy. So, even if the learners do not have any computer experience, they can use the system very easily. The hardware architecture of the system is shown in Figure 2.

3.2 The software architecture

The software architecture of the system is shown in Figure3. There are 4 databases, which store text
material, voice material, picture material, and the usage behavior record of learners, respectively. The text material database stores the English words, phrases, and sentences of all objects. The voice material database stores the English and Chinese pronunciation of all teaching material. The picture material database stores the animation and image of all objects. The learners’ portfolio database records the behavior operation process, for example, the setting of personalized preference, the time of all objects scanned, the time of the function button clicked.

![Figure 3. The software architecture.](image)

3.3 The snapshots when the system was executed

The operation of the system is very simple. After starting the system, learners just need to scan the real objects with smart phones or tablets as shown in Figure 4.

![Figure 4. Scan the actual scene objects with a tablet.](image)

The system provides 3 learning scopes: phrases, sentences, related words. In the learning process, learners can turn on or off these functions with their own preferences. In addition, there are 2 learning modes can be selected. One is the easy mode and the other is the advanced mode. These functions are for customized design and as shown in Figure 5.
When learners successfully scan an object, the screen will show the main teaching material (for example, a pen) and the teaching material of related word (for example, an eraser). A flashing light will appear around the two teaching materials to enhance the display, so that the learners can clearly know where the operable objects are. After the learners click the object that want to learn (For example, the eraser), the corresponding function buttons will appear, as shown in Figure 6. Such design is to meet the "Spatial Contiguity Principle" and "Temporal Contiguity Principle" of 12 multimedia design principles proposed by Mayer (2009). The design allows learners can simultaneously observe the real learning object and corresponding virtual teaching material. Thus, the best learning effectiveness can be obtained.

When learners click the function button of "words", the system will first split the word into letters and read out each letter. Meanwhile, the screen also appears the corresponding letters in marquee style. Then, the word will be read out once. Then, the Chinese meaning will be explained using Chinese voice, as shown in Figure 7. Such design is to meet the "Modality Principle" of 12 multimedia design principles proposed by Mayer (2009). The design allows the learners can use auditory and visual multisensory to receive a single message. Thus, the best learning outcomes can be achieved.
If learners click the function button of "phrases", the teaching material of phrases will appear on the screen. Then, the phrase will be read out using English and the Chinese meaning will be explained using Chinese voice. In addition, the corresponding animation will appear. The system will give simple phrase in the easy mode (as shown in Figure 8) and give more difficult phrase in the advanced mode (as shown in Figure 9).
If learners click the function button of "sentences", the teaching material of sentences will appear on the screen. Then, the sentence will be read out using English and the Chinese meaning will be explained using Chinese voice. In addition, the corresponding animation will appear. According the easy or advanced mode, the degree of difficulty of the sentence is also different, as shown in Figure 10.

![Figure 10. The function of “sentences” button.](image)

4. Conclusions and Future Works

This study proposed a customized AR English learning system. The system provided 3 learning scopes (phrases, sentences, and related words) and 2 modes (the easy mode and the advanced mode). By AR technology and customized design, the learning motivation and effectiveness of learners are expected to be enhanced.

In the future, we will invite 2 teachers in an elementary school in the central region of Taiwan to provide 100 English vocabularies suitable for grade 5 students and to conduct experimental teaching. The experimental subjects are 100 students of 4 classes. All the operation process will be recorded in database, so that further personalized preference analysis can be conducted.

Acknowledgements

This study is supported in part by the National Science Council of the Republic of China under Contract No. MOST 103-2511-S-275-002-MY2.

References


Development of a Multi-Device Data Structures Course Item Bank Practice System with Self-Regulated Learning Strategy on Bloom's Taxonomy of Educational Objectives

Gwo-Haur HWANG*\textsuperscript{a}, Jing-Fang CHEN\textsuperscript{a}, Yu-Ting SHIH\textsuperscript{a}, Yong-Sheng JHANG\textsuperscript{a}, Yi-Xuan LIN\textsuperscript{a} & Yu-Syuan WANG\textsuperscript{b}

\textsuperscript{a} Department of Information Networking and System Administration, Ling Tung University, Taiwan
\textsuperscript{b} Department of Mathematics and Information Education, National Taipei University of Education, Taiwan
*ghhwang@teamail.ltu.edu.tw

Abstract: With the rapid development of information technology, traditional paper-and-pencil testing is lacked for immediate analysis and feedback. Thus, it has been replaced by computerized item bank practice systems. Because of the popularity of wireless networks and multi-devices, developing an item bank practice system which can support on multi-device has become a new trend gradually. This makes learning become more convenient. In this paper, we classify items by Bloom's taxonomy of educational objectives, and found the weaknesses of learners in the practice in order to improve the weaknesses. At the same time, we use self-regulated learning strategy in this system, so that learners can set their learning goals. In this system, it provides self-monitoring, standard setting, evaluative judgment, self-appraisal, and affective self-reaction, so that learners can learn in the best environment. In the future, we will combine this system with data structures course and hope to improve learners’ learning motivation and effectiveness.

Keywords: Multi-device system, data structures, item bank practice system, self-regulated learning, Bloom's taxonomy of educational objectives

1. Introduction

The traditional practice method is a paper-and-pencil testing, and it is replaced by computers. When learners encounter problems, they need teachers help, otherwise students cannot get the correct information immediately. With the rapid development of information technology and internet, computer online practice has become the new trend gradually. Tu (2003) developed an on-line assessment system, which combined a natural sciences course. The system can be repeated practicing and provide the feedback immediately. In addition, Lee, Tseng and Tsai (2003) and Chen, Chang and
Wang (2008) mentioned, because of the popularity of wireless networks and multi-devices, learners can learn online via multi-device to improve their learning effectiveness. Therefore, multi-device item banks practice come into being. For example, Kung, Huang and Chung (2007) developed a multi-device learning assessment system which combined class b of computer software application technicians. The results of their study showed that the system improved learners’ learning effectiveness. However, these papers only developed multi-device item bank practice systems, but did not enhance learners’ weaknesses. Therefore, we use the Bloom's taxonomy of educational objectives to analyze the weaknesses of learners.

Bloom, Engelahar, Frust, Hill and Krathwohl (1956) proposed the “Bloom's Taxonomy of Educational Objectives”, which was used by many educators and updated to the new version (Anderson & Krathwohl, 2001). The Bloom's Taxonomy of Educational Objectives has good effectiveness on designing items and diagnosing learners’ weakness (Crowe, Dirks, & Wenderoth, 2008; Hwang, Chen, Loe, & Huang, 2013). Therefore, we import the taxonomy and wish let learners can understand their weaknesses, so that they can adjust the learning goal by themselves and improve their learning motivation. However, initiative is a critical factor for learners’ learning effectiveness (Govaere Jan, de Kruif, & Valcke, 2012). Thus, we also import the self-regulated learning strategy in this system to improve learners’ initiative.

Learners can set a goal to carry on self-monitoring, standard setting, evaluative judgment, self-appraisal, and affective self-reaction, which is spirit of self-regulated learning (Bandura, 1991), i.e., learners can learn through setting and adjusting goals. We hope to improve learners’ learning motivation and effectiveness via self-regulated learning. However, if learners only depend on the behavior of initiative investment and self-monitoring without the effective adjusting strategies, learners’ learning effectiveness cannot be improved (Zimmerman & Kitsantas, 1997). Therefore, the effective adjusting strategy is a critical factor. Wang (2011) indicated that adding the self-adjust learning mechanism into formative assessment and using the Peer-Driven Assessment Module (PDA) strategy can lead to learners’ learning motivation. The formative assessment can be learners’ learning basis. Learners can adjust their goals repeatedly. The experimental results indicated that learners who use the PDA-WATA (Web-based Assessment and Test Analysis) system are better than learners who use the N (None)-WATA system on initiative and learning effectiveness. Therefore, we also provide all learners complete rate ranking, which can motivate learners via comparing.

In the past, about the self-regulated learning, many scholars found that learners can achieve good learning effectiveness via the self-regulated learning strategy in mathematics (Hackett & Betz, 1989; Malpass, O'Neil, & Hocevar, 1999; Pajares & Miller, 1994; Parker, Marsh, Ciarrochi, Marshall, & Abduljabbar, 2014) and natural sciences (Betz & Hackett, 1983; Britner & Pajares, 2006; Chen & Usher, 2013) domains.

Therefore, according to above mentions, we develop a multi-device data structures course item bank practice system with self-regulated learning strategy on Bloom's taxonomy of educational objectives. The Bloom's taxonomy of educational objectives and the self-regulated learning strategy can help learners understanding their weaknesses to practice and give the appropriate feedback. At the same time, learners can use multi-device learning tools to learn without any limitation of space and time. We expect that learners can improve their learning motivation and effectiveness via this system.
2. Literature

2.1 Item Bank Practice System

Bunderson, Inouye, and Olsen (1989) indicated that computerized testing includes four steps of development: (1) computerized testing: the traditional paper-and-pencil testing is replaced by using computer to practice; (2) computerized adaptive testing: item response theory (IRT) is applied to conduct the computerized adaptive test (CAT); (3) continuous measurement: history records are added; (4) intelligent measurement: expert systems with artificial intelligence provide learning suggestion to learners. The first three steps are lacked for immediate feedback to carry on effective improvement in traditional teaching. However, with the popularity of internet, item bank practice systems are changed from single-computer practice to computer-online practice (Hwang, Chen, Huang, & Loe, 2013; Tu, 2003; Yo, Jan, & Li, 2011). It is an issue how to improve learners’ learning effectiveness by item bank practice systems. In this paper, the system combines the history records of the third step and the expert system of the fourth step, and assists learners to proceed to effective learning. On the other hand, multi-device learning tools make learners learn conveniently.

2.2 Multi-Device Learning

Gay, Stefanone, Grace-Martin and Hembrooke (2001) used laptops as learning tools on the communication course and the computer science course. Using of wireless Internet technology, learners can discuss course anywhere. Thus, learning by mobile device is feasible. Guerreroa, Ochoaa and Collazosb (2010) built a learning system. In this system, learners can use PDAs to carry on the grammar practice online. Learners are divided into some groups can discuss courses immediately. Teachers can see the learners’ answers and comments and reduce the time of marking learners’ homework. The experimental indicated that more than 70% of the learners who used the system can improve the language grammar ability, and more than 86% of the learners thought the system is operated easily. The system can not only improve learners’ learning motivation and effectiveness, but also reduce the load for teachers. With the rapid development of multi-device learning tools (smart phones, tablets, notebooks, PDAs and computers), learners have more choices. In the empirical researches, Chen et al. (2008) and Hwang et al. (2013) had established two learning websites which allowing learners to use multi-device learning tools for learning without limitation all the time. At the same time, the learners’ portfolios are recorded on the websites, so that teachers can see the learning situation of learners. The experimental results indicated that the webs can improve the performance of learners. As the above literatures mentioned, we can know that using multi-device learning tools will make learners learn conveniently. However, it is important to this paper how to recognize learners for items understanding. Thus, we use the Bloom’s taxonomy of educational objectives to find the learners’ weaknesses and improve them.

2.3 Bloom's Taxonomy of Educational Objectives

Bloom et al. (1956) propose cognitive domain taxonomy of educational objectives as “Bloom's taxonomy of educational objectives”, which includes six classes. From easy to difficult, the six classes
included knowledge, comprehension, application, analysis, synthesis and evaluation. With the rapid
development of educational psychology, Anderson and Krathwohl (2001) corrected the Bloom's
taxonomy of educational objectives. This taxonomy includes knowledge dimension and cognitive
process dimension. Knowledge dimension is subdivided into factual knowledge, conceptual
knowledge, procedural knowledge and metacognitive knowledge. Cognitive process dimension is
divided into six categories, i.e., remember, understand, apply, analyze, evaluate and create. Therefore,
many scholars have used this taxonomy to classify items. Lan and Chern (2010) classified the university
entrance exams in English reading questions by the Bloom's taxonomy of educational objectives, which
can let teachers catch the emphasis to carry on teaching easily and let learners more understand the topic
and the problem-solving elements. Therefore, in this paper, we use remember, understand, apply and
analyze of cognitive process dimension to classify items of the data structures course. We hope to
analyze learners’ weaknesses effectively and let learners improve their learning motivation and
effectiveness by the self-regulated learning strategy.

2.4 Self-Regulated Learning

Bandura (1977) indicated the self-regulated learning conception as self-efficacy includes the efficacy
expectation and the outcome expectation. The efficacy expectation means to set goals and to do
self-evaluation for these goals. The outcome expectation means that, if learners think that their ability
can’t reach their goals, they cannot keep their learning motivation. Therefore, we think that self-efficacy
(self-regulated learning) is important to impact learners’ learning motivation. Subsequently, Schunk
and Zimmerman (1994) indicated that the self-regulated learning includes four steps, i.e.,
“self-evaluation and self-monitoring”, “goal setting and strategy planning”, “using and monitoring of
strategies” and “monitoring of results of strategies”. Learners can learn according to the learning pace
by themselves via these four steps. In summary, all scholars think that setting goal and self-evaluation in
self-regulated learning are very important. And Schunk and Zimmerman also proposed “using and
monitoring of strategies” and “monitoring of results of strategies”. In the self-regulated learning
environment, learners can not only set goals and carry on self-evaluation, but also they can carry out
self-adjustment via the learning results. In addition, Multon, Brown and Lent (1991) aim related 39
papers of self-regulated learning to carry on integrated analysis in the past. The results of their study
showed that the self-regulated learning can improve learners learning effectiveness in different subject
areas, and different assessment methods. Therefore, we develop a multi-device data structures course
item bank practice system with self-regulated learning strategy on Bloom's taxonomy of educational
objectives. We hope this system can improve learners’ motivation and effectiveness of learning.

3. Development of Our System

3.1 System Architecture

In this paper, we use Windows Server 2008 to set up an Internet Information Services (IIS) web server
and a Microsoft SQL database server, and develop all modules by Visual Studio 2010 ASP.NET C#
language. This system is divided into two interfaces, i.e., the learner interface and the teacher interface.
The learner interface includes three modules, i.e., “goal management module”, “item bank practicing module” and “personal learning portfolio inquiry module”. Learners can carry on operating of three modules with multi-device learning tools, e.g., cellphones, tablets and computers. In order to implement the self-regulated learning strategy, “goal management module” can provide learners setting goals which include selecting range, Bloom's Taxonomy of Educational Objectives (remember, understand, apply and analyze), and passing score. “Item bank practicing module” can provide learners to carry on practice via selected range by learners, and it is based on the Bloom's taxonomy of educational objectives to display the weaknesses of learners. Learners can understand their weaknesses and adjust the goals in “goal management module”. “Personal learning portfolio inquiry module” can provide learners to inquiry the practice records in the past. On the other hand, the teacher interface includes three modules, i.e., “learners’ basic information management module”, “item bank management module” and “learning portfolio management module”. Teachers can operate these three modules by computers. “Learners’ basic information management module” can provide teachers to modify learners’ basic information. “Item bank management module” can provide teachers to modify items and set the Bloom's Taxonomy of Educational Objectives (remember, understand, apply and analyze) of items. “Learning portfolio management module” can provide teachers to inquire the practice records of all learners’ learning situation. All modules can access “learners’ basic information database”, “item bank database” and “learning portfolio database”, as shown in Figure 1.

![Figure 1. System architecture.](image)

3.2 Operating interface

Learners can watch the currently complete rate ranking on the login snapshot of the system, as shown in Figure 2. After learners sign in the system, they can see the three buttons (“system homepage”, “goal management and item bank practicing” and “personal learning portfolio inquiry”) on left side, as shown in Figure 3.
When learners click the “goal management and item bank practicing” button, and enter “goal management module”, they can select chapters to practice by themselves. The items of every chapter are classified according to Bloom’s Taxonomy of Educational Objectives (remember, understand, apply and analyze), and learners can do practice for their selected types. In addition, the system also provides learners a setting goal function. Learners can set the passing score of the current practice, as shown in Figure 4. Subsequently, learners can carry on practice in “item bank practicing module”. The practice interface of learners is shown in Figure 5.
End of the practice, the interface will change to the “goal management module”. Learners can see the correct rate, the courses complete rate and the feedback of practicing. There are two buttons at the bottom of “goal management module”, i.e., the “reset goal” button and the “practice again” button, are provided learners resetting goal to practice or practice again of the same goal, as shown in Figure 6, 7.

When learners click the “personal learning portfolio inquiry” button and enter the “personal learning portfolio inquiry module”, they can see the practice records quickly via the drop-down list above the interface. In the drop-down list, the red items mean that the practice is no-passed, and the black items mean passed, as shown in Figure 8. Learners can inquiry practice records after practicing, and they can watch all practice items or only wrong items, so that they can review and correct. Learners can also watch items quickly via the drop-down list on the right hand of the interface. In the drop-down list, the red items mean that learners answer incorrectly, and the black items mean that leaners answer correctly right, as shown in Figure 9.

In the teacher interface of “learners’ basic information module”, which is provided teachers to inquiry the basic information of all learners. In “item bank management module”, teachers can click the
“item bank management” button and carry on the management of items, as shown in Figure 10. In “learning portfolio management module”, teachers can watch all learners’ portfolios.

![Image](image1.png)

**Figure 10.** Item bank management.

4. Conclusion and future works

4.1 Conclusion

In this paper, we develop a multi-device data structures course item bank practice system with self-regulated learning strategy on Bloom's taxonomy of educational objectives. Learners can use multi-device to practice item bank. At the same time, they can set a goal to learn and budget time by themselves. In addition, every question in the item bank is set from a Data Structures teacher, and they are classified according to remember, understand, apply, and analyze. Therefore, this system can make learners clearly know what their weaknesses are. We expect learners can improve their motivation and effectiveness of learning via this system.

4.2 Future work

In the future, we will carry on an empirical research with two classes of attending data structure course which combine this system. The participants are about 99 learners in two classes. One class will be the experiment group which will use the self-regulated strategy. Learners of the experiment group can set a goal includes selecting range, Bloom's Taxonomy of Educational Objectives (remember, understand, apply and analyze), and passing score by themselves. The other class will be the control group which will use the non-self-regulated strategy. Learners of the control group will go by what the homework request of teacher to do practice. If they won't complete in time, they can't do practice. Their passing score will be set to 75. Two classes will proceed with pre-test, post-test and questionnaire (learning effectiveness, learning motivation, learning attitude, learning satisfaction and cognitive load). After the experiment, we will analyze to learners' learning motivation and learning effectiveness between the experimental group and the control group, as shown in Figure 11.
Figure 11. The flowchart of the experiment.

Acknowledgements

This study is sponsored by the Ministry of Science and Technology in Taiwan under the contracts no. MOST 103-2511-S-275-002-MY2.

References


Students' Self-efficacy and Acceptance toward Context-Aware Ubiquitous Learning in Biology Education: A Case of Photosynthesis in Plant

Chuntanet NASAROa & Niwat SRISAWASDb*

aScience and Technology Education Program, Faculty of Education, Khon Kaen University, Khon Kaen, Thailand
bDivision of Science, Mathematics, and Technology Education, Faculty of Education, Khon Kaen University, Khon Kaen, Thailand

*niwsri@kku.ac.th

Abstract: Recent progression in wireless and sensor technologies has led to a new development of learning environment, called context-aware ubiquitous learning environment, which is able to transform students’ learning experience into more authenticity. With the benefit of the context-aware u-learning environment, this study aims to investigate impact of sensor-based laboratory learning in biology, incorporated with predict-observe-explain (POE) pedagogy on students’ perceptions of self-efficacy, perceived usefulness, and perceived ease of use. The 21-item Likert-scale questionnaire was administered to 44 eleventh-grade students both before and after the intervention. The results indicated that the POE-based ubiquitous learning in the real context of plant photosynthesis could promote students’ self-efficacy and perceived ease of use on the learning. However, there was no significant effect on perceived usefulness of the intervention. As such, this study also provided the discussion for promoting the students’ perceived usefulness after participating in the POE-based tablet.

Keywords: Ubiquitous learning, POE, ubiquitous computing, biology education, wireless networks

1. Introduction

Recent development in science and technology applications have influenced on learning process. In 21st century learning, computer and wireless network technologies have greatly affected the delivery of learning and capacitated people to convenient. The widespread of mobile devices, such as Tablet, PC, PDA and smart phone, has transformed learning modes from e-learning to m-learning (Minami, Morikawa and Aoyama, 2004). These technologies provide learning opportunity to everyone, anywhere and anytime. Mobile technology provides opportunities to support science learning both inside and outside classroom. With mobile technology, the learning environment can go with the student to outdoor, the laboratory and other beyond.

Chen, Chang and Wang (2008) suggested that ubiquitous learning is context sensitive anytime, anywhere learning using ubiquitous devices such as, Tablet, Web, PDA’s in indoor, outdoor, individual and groups. Schroeder and Haskell (2011) described u-learning as social media plus mobile learning. Some of these characteristics are applicable to here and now learning (Huang, Chiu, Liu and Chen, 2011): (i) Hastening of learning need (on time suitable) (ii) Initiative of knowledge occupancy (providing information to students immediately) (iii) Situation of learning activity (flow of everyday activities) (iv) Context awareness (interaction controlled by context such as location, time, activity etc) (v) Self-regulated learning (students control their learning process)

Cause many parts of science topics difficult to understanding and learning in many area teaching by textbook which is low motivation and better to learn which should to learn in real phenomena. The prevalence of outdoor education has increased considerably in recent years (Bloom, Holden, Sawey and Weinburgh, 2010). Students can use those outdoor learning experiences to understand and establish new knowledge and concepts regarding the topic being studied (Auer, 2008, Upadhyay and DeFranco, 2008). On the other hand, Teachers can incorporate knowledge regarding
ecology that students gained through outdoor learning into formal classroom instruction to improve students' perception (Eick, 2012). Thus recent u-learning should be appropriately used to support authentic learning in real phenomena as science subject.

2. Purpose of the Study

Based on the abovementioned rational, this study aims to investigate students' perceptions delivered in sensor-based laboratory learning environment incorporated predict-observe-explain (POE) pedagogy for biology learning of photosynthesis. Specifically, the following questions were answered:

- Do the students engaged in sensor-based laboratory learning environment incorporated predict-observe-explain (POE) perform significantly better by students’ perceptions of self-efficacy?
- Do the students engaged in sensor-based laboratory learning environment incorporated predict-observe-explain (POE) perform significantly better by students’ perceptions of perceived usefulness?
- Do the students engaged in sensor-based laboratory learning environment incorporated predict-observe-explain (POE) perform significantly better by students’ perceptions of perceived ease of use?

3. Literature review

3.1 A Context-aware Ubiquitous Learning

Recent progress in wireless and sensor technologies has lead to a new development of learning environments, called context-aware ubiquitous learning environment, which is able to sense the situation of learners and provide adaptive supports based on radio-frequency identification (RFID), wireless network, embedded handheld device, and database technologies. Many researchers have been investigating the development of such new learning environments. For example, a context-aware mobile learning system was used as a sensing device for nursing training courses and they found that students' learning outcomes were notably improved by utilizing the mobile learning system (Chen and Huang, 2012). Shih, Chuang and Hwang (2010) study with fifth grade students with the inquiry-based mobile learning system. They investigate by pre- and post-test with observations and interview focus group. The finding showed significant positive results for students' learning. In an addition, e-library activity worksheets were developed by Hung, Lin and Hwang (2010) that helped students focus their outside ecology observation tasks, and results indicated that most students demonstrated substantial improvements and extended their inquiry skill. These evidence indicated potentials of the novel learning environment of context-aware ubiquitous learning in teaching and learning.

3.2 Inquiry-based Learning

Kuhn, Black, Keselman and Kaplan (2000) descript that inquiry-based learning is one of primarily pedagogical based on the investigation of questions. By the process of investigation and collection of science data, inquiry activities provide a valuable context for learners to acquire, clarify, and apply an understanding of science concepts (Edelson, Gordin and Pea, 1999). Furthermore, many researchers try to develop learners' investigation skills, data analysis and critical thinking using inquiry-based learning. They adopt activities related to the natural world to allow students to observe events and objects in the physical world from various facets, and to develop an understanding of how scientists explore the natural world (Hmelo-Silver, Duncan and Chinn, 2007). The advantages of inquiry learning are that it can lengthen the retention period of new knowledge, increase problem solving flexibility and creativity, and increase student learning motivation (Lord and Orkwiszewski, 2006). When inquiry learning is used in science subjects, it shows great potential for increasing students' understanding of scientific knowledge and their engagement in science.
POE strategy is a constructivist-based pedagogy and many researchers employed to facilitate learner’s conceptual change process in inquiry-based learning. This kind of settings may provide a powerful learning environment for students where they have opportunities to construct scientific conceptual understanding that is durable over time (de Jong, 2005). POE can provide students to work on tasks collaboratively in group. Thus, it encourages a cooperative learning environment where students can share their knowledge and discuss with others in their group (Küçüközer, 2008, Tao and Gunstone, 1999).

4. Method

4.1 Study Participants

A total of 44 student-respondents in their eleventh grade, age ranging from 16 to 17 years in a local public school at the northeastern region of Thailand participated in this study. They were attending a biology course for basic education level. They have no experience yet using sensor-based laboratory in biology learning. This implied that they are heterogeneous before interacting with the experimental study.

4.2 Learning Materials and Activity

To engage student into context-aware ubiquitous learning, this study employed wireless microcomputer-based laboratory by Vernier and software technology. Vernier LabQuest-2 is a standalone interface used to collect sensor data with its built-in graphing and analysis application. The large, high-resolution touch screen makes it easy and intuitive to collect, analyze, and share data from experiments in class, field, and anyplace. Its wireless connectivity encourages students’ collaboration and personalized learning. For this study, students were provided opportunity to conduct an investigation of photosynthesis in plant. With wireless features of the Vernier laboratory, the researchers designed students’ laboratory experience with measurement of rate of oxygen (O$_2$) and carbon dioxide (CO$_2$) released by plants in the real context by gas sensors. The Vernier CO$_2$ and O$_2$ gas sensors used to measure gaseous carbon dioxide by monitoring the amount of infrared radiation absorbed by carbon dioxide molecules, and gaseous oxygen levels in a variety of environment, respectively. With the use of wireless LabQuest-2 data logger, experimental data were obtained, processed, and then shared to tablets in classroom via a server, as seen in Figure 1.

![Figure 1. Materials of tablet-based laboratory learning.](image)

To promote students’ self-efficacy and acceptance toward the learning intervention, constructivist POE (predict-observe-explain) sequential learning steps were designed to foster biology learning by inquiry. In this study, instructor used 100 minutes class session to provide the u-learning experience for biology class. Student took overall 30 minutes for pre-test and post-test, and another 70 minutes for their POE learning process. Before performing the learning process, the instructor gave an orientation for working with the Vernier laboratory in 5 minutes and then 65 minutes for prediction (5 minutes), observation (30 minutes), and explanation (30 minutes). Figure 2 represented the prediction step based on an open-ended inquiry question provided by instructor, “what will happen if we measure rate of CO$_2$ and O$_2$ from plans outside the classroom?” Then, students predict the graph of rate of O$_2$
and CO\textsubscript{2} when plant live in light and dark area on a work sheet. This warm-up activity was designed to provide the students with basic knowledge, and stimulate their motivation to learn science.

![Image](image1.png)

**Figure 2.** Prediction Stage. Students drawing graph on work sheet.

In the next learning step, a volunteer group of students was setting up by instructor and assign them to investigate rate of photosynthesis of plants using the Vernier gas sensor and LabQuest-2 data logger at the field for 30 minutes, regarding the variations in amount of carbon dioxide and oxygen. Figure 3 illustrates the observation activity in the real context.

![Image](image2.png)

**Figure 3.** Observe Stage. Students observe by authentic measure rate of plants release and absorb O\textsubscript{2} and CO\textsubscript{2} in real area and process result by Labquest2 and share data on tablet in classroom by wireless connection.

At the same time with the observation, another student viewed the real-time experimental data of rate of plant photosynthesis through tablet connected wireless internet system in classroom. Figure 4 illustrates the use of tablet to monitor results of photosynthesis experimentation.

![Image](image3.png)

**Figure 4.** Observe Stage. The results were shared on tablet in classroom by wireless connection. Students compared their predictions and real observe.
For the last learning step of explanation, the volunteer group of students came back to the classroom and then the instructor conducted a forum discussion with peers to collaborative explain the plant photosynthesis phenomenon, emphasizing comparison of their prediction and explanation (See Figure 5).

![Figure 5. Explain Stage. Students discuss and explain their prediction and real observe.](image)

### 4.3 Instruments

A 21-item Likert-scale questionnaire was developed to use in this study for examining students’ self-efficacy, perceived usefulness, and perceived ease of use toward the context-aware ubiquitous learning of plant photosynthesis. There were eight items of self-efficacy obtained from Wang and Hwang (2012), and six and seven items of perceived usefulness and perceived ease of use, respectively, from Hwang, Yang and Wang (2013). To develop a Thai version of the questionnaire, the original English version (See Table 1) was translated identically in Thai language. One expert was recruited to identify communication validity of the items. On each item, respondents were assigned to rate how much the respondent agree with into five scale, from 1-strongly disagree to 5-strongly agree). The reliability for self-efficacy, perceived usefulness, perceived ease of use, overall items was 0.76, 0.47, 0.73, and 0.85 (N=40), respectively.

<table>
<thead>
<tr>
<th>Scale</th>
<th>Description of the scale</th>
<th>Example of items</th>
</tr>
</thead>
<tbody>
<tr>
<td>Self-efficacy</td>
<td>Self-efficacy, or students’ beliefs regarding their capability to execute actions necessary to achieve designated outcomes</td>
<td>. I am confident that I can learn the basic concepts well of this work.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>. I am confident that I can finish this work well.</td>
</tr>
<tr>
<td>Acceptance of the technology or learning approach</td>
<td>To better understand the students’ perceptions of tablet-based laboratory learning approach, the students’ ratings for the “perceived usefulness,” and “perceived ease of use”</td>
<td>Perceived Usefulness</td>
</tr>
<tr>
<td></td>
<td></td>
<td>. The learning mechanisms provided by the learning system smoothed the learning process.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>. The learning approach is more useful than the conventional computer-assisted learning approaches.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Perceived Ease of use</td>
</tr>
</tbody>
</table>
4.4 Data collection and Analysis

Figure 6 shows the procedure of the experiment. Before the learning activity, the students took the pre-test questionnaire. During the learning activity, stage 1 teacher provide inquiry question “what will happen if we measure rate of CO₂ and O₂ from plans now?” that students predict graph of result on activity sheet in each group (15 minutes). Then stage 2, students observe in field site by material learning (Vernier laboratory). The result will be show on tablet in each group inside classroom (30 minutes). Stage 3 student discuss in their group, compare the result of prediction and observation and explain (30 minutes)

After the learning activity, the students took the post-test questionnaire for comparing the pretest and the improvements learning.

![Figure 6. Diagram of experiment design.](image)

The statistical data analysis techniques selected for this study were pair sample t-test. When there are measurement and nominal variables. The paired t-test in SPSS is used to compare the values of means from two related samples in a before and after learning.

5. Results and Discussion

In order to explore effects of the sensor-based laboratory learning environment incorporated predict-observe-explain (POE) for biology learning of photosynthesis, Table 2 shows results on students' self-efficacy, perceived usefulness, and perceived ease of use.

<table>
<thead>
<tr>
<th>Variable</th>
<th>N</th>
<th>Mean</th>
<th>SD</th>
<th>t</th>
</tr>
</thead>
<tbody>
<tr>
<td>Self-efficacy</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pre-test</td>
<td>44</td>
<td>27.41</td>
<td>.555</td>
<td>2.846*</td>
</tr>
<tr>
<td>Post-test</td>
<td>44</td>
<td>29.80</td>
<td>.665</td>
<td></td>
</tr>
</tbody>
</table>
The results show that student' self-efficacy and perceived ease of use delivered in the context-aware u-learning were statistically significant difference between pretest and posttest (t = 2.846, p<0.05 and t = 4.310, p<0.05, respectively). The result indicated also that the students’ perceptions on self-efficacy and perceived ease of use showed significant improvement after participating with the context-aware u-learning. These means that learning with the sensor-based laboratory learning in biology, incorporated with predict-observe-explain (POE) pedagogy was not difficult and they were able to perform the investigation of plant photosynthesis with peers to accomplish the predetermined learning task in biology class. The result is consistent with the research findings that students with a strong sense of efficacy are more likely to challenge themselves with difficult tasks and be intrinsically motivated (Margolis and McCabe, 2006). In an addition, this finding consistent with Shih, Chuang and Hwang (2010), which reports that students felt more liberating and relaxing than learning in the classroom with the inquiry-based learning experience in field trip. Also, Hwang, Yang and Wang (2013) reported that students showed positive feedback after participating with the context-aware u-learning. However, the result indicated that there was no significant difference on students' perceived usefulness between pretest and posttest (t = 1.335, p<0.05). This result expressed that the students did not perceived the usefulness of the sensor-based laboratory learning incorporated with predict-observe-explain (POE) pedagogy for their biology learning. The reason might be that they did not familiar with the use of tablet technology in science learning and never use sensor-based laboratory learning for doing science before. Therefore, they might have no idea how this kind of learning environment would be benefits to them. According with Hwang, Tsai and Yang, (2008) descript that there are several levels of individualized guidance, to support learning with a context-aware u-learning environment, for naive learners, adaptive supports and guidance for real-world operations or observations should be provided for learners who have different backgrounds and experiences.

6. Conclusions

Although ubiquitous learning or context-aware learning environments is still available in Thailand, this paper is an effort to design, develop, and implement context-aware ubiquitous learning experience for biology learning in school science. This study demonstrates how instructions using context-aware ubiquitous learning with a sequential learning process of predict-observe-explain promoted students' perceptions on self-efficacy and perceived ease of use in teaching and learning about photosynthesis of plant phenomena. The results suggested that student can perform science investigation with sensor-based learning technology effectively. However, to promote students’ perceived usefulness of the ubiquitous learning experience, the intervention in this study may need a revise to contribute fostering their perception of usefulness following Tsai, Tsai and Hwang (2011), which explored students’ conceptions of context-aware ubiquitous learning and they found that students perceived the learning environment as followings: application of technology in the learning process; a convenient way to attain information to achieve their goals; a timely guide to apply the mobile devices to provide directions for learning; increase of knowledge; and a way of allowing them to engage in inquiry practices, such as allowing open-ended exploration for the learning topic.

Acknowledgements

This work was financially supported by the Institute for the Promotion of Teaching Science and Technology (IPST) and Graduate School, Khon Kaen University. The author would like to express
gratefully acknowledge to Faculty of Education, Khon Kean University, for supporting this contribution.

References


How Self-Efficacy Affects Students’ Performance and Pace in Self-Directed Learning with ICT

Andrew C.-C. LAO*, C.-L. HUANG*, O. KU & T.-W. CHAN*

*Graduate Institute of Network Learning Technology, National Central University, Taiwan
bResearch Center for Science and Technology for Learning
*ccl.andrew@gmail.com

Abstract: Due to the fact that not only most Self-Directed Learning (SDL) studies remains discussed at the undergraduate or adult level, but also ICT in education has been positively proved as a compelling factor for the children, a SDL with ICT scenario is described in this paper. In SDL, with or without the assistance or guidance from the classroom teacher, students learn by their own during the learning activity, in which the students will identify their needs, set their goals, take learning missions, and reflect their learning outcome by reviewing their finished learning missions and performance. However, although students’ self-efficacy affects how the students own the perception on self-management or goal accomplishments (Ormrod, 2006), limited discussion is found for exploring self-efficacy in SDL with ICT. Therefore, in this study, students are pre-categorized into high and low self-efficacy groups for exploring the effect of self-efficacy for SDL with ICT. As a result, the result shows that students’ self-efficacy might not be a crucial factor that affect students’ performance or pace, with only a slightly higher gain score on performance is found in this study.

Keywords: Self-directed learning, SDL, self-efficacy, ICT

1. Introduction

Self-Directed Learning (SDL) is believed as an essential andragogy in promoting students’ individualization (Knowles, 1975). However, the original design of SDL lies within adults or adolescents (Gibbons, 2003; Knowles, Holton, Swanson, 2012), and consequently limited discussions were found at the elementary level. The reasons to the phenomenon, on the one hand, might be due to the children’s maturity or cognitive engagement problems (Guthrie, 2004; Pressley, 2006; Taylor, Pearson, Peterson, & Rodriguez, 2003). On the other hand, the rationale for most of the public elementary curriculum restricts the flexibility for individual students’ development, even though every student holds different capability in learning. Nevertheless, with regards to the aforementioned problems, compare to children in decades before, most of the children nowadays, could be more independent from the parents or teachers, and they could be more capable to learn individually (Glaubman, Glaubman, & Ofir, 1997; Philips & Stipek, 1993). Therefore, an increasing number of researchers pay attention to the development of SDL at the elementary level (Teo, Tan, Lee, Chai, Koh, Chen et al., 2010), but only a few studies discussed the ICT adoption in public schools with SDL. Followed by the master plan from the government, Teo et al. explored the possibilities of SDL with technology readiness in terms of learning goals, resource allocation, planning, monitoring, and reassessment of learning strategies.

Moreover, most studies have empirically proved the positive effects for students’ cognition or affection by using ICT in education. In other words, ICT shows potential in public classrooms for education. Therefore, it could be concluded that the computer technology not only allowed students to be more engaged, but also led to a significant improvement on learning performance. In addition, since students’ self-efficacy affects how the students manage to finish the learning goals for the learning activity (Kim, 2009), the effect for the self-efficacy in SDL should be considered. Hence, with regards to the importance of SDL and the positive effect of ICT, this study designs a SDL scenario that based on
the rationale by various researchers related to SDL, and provides a preliminary analysis for the effect of different levels of self-efficacy that related to students’ performance and pace.

2. Related Work

2.1 The research related to Self-Directed Learning

Self-Directed Learning (SDL) helps promote students' individuality in learning and illustrate how the students' intention and action for learning individually (Tough, 1971; Knowles, 1975; Knowles, Holton & Swanson, 2012). The original idea for SDL could be traced to Tough’s study, which was called “self-planned learning”. In Tough’s study, he found that most students or learners frequently applied the 3Ws (what, where, and when) in learning. The 3Ws could be considered as guidance for designing the SDL scenario, where students should be able to determine what to learn, where to start, and when to finish the learning activity. Besides, researchers like Tough and Knowles both shared the similar concept in SDL, but Knowles had later popularized the term “Self-Directed Learning” and offered many learning resources for students and teachers to apply SDL in classrooms. For example, Knowles and his colleagues believed that, in SDL, students are responsible for their decisions in the learning activity, while students would set and develop their personal goals, revise and reflect their own pace or learning experience. In addition, Knowles (1975) also proposed six steps for the implementation of SDL in classrooms:

- Setting the classroom environment
- Learning needs
- Learning goals
- Identifying learning resources
- Applying suitable learning strategies
- Evaluating the learning outcome

The six steps presented above played an essential role for the adoption of SDL in the classrooms, because the steps emphasized on the transformation of the classroom learning from teacher-centered to student-centered, and focused on not only the knowledge acquisition, but also the personal reflection for the learning outcome. Therefore, followed by the SDL concept by Tough or Knowles, a researcher such as Gibbons (1994, 2002) provided specific criteria for SDL in practice. More specifically, to facilitate SDL in classrooms, from Gibbons’ two studies, he suggested various criteria from two perspectives that are related to the development of the SDL framework, in which adolescents learned under the guidance by the school teachers. (See Table 1).

Table 1: Gibbons’ two studies for facilitating SDL in classrooms.

<table>
<thead>
<tr>
<th>Teacher’s perspectives (Gibbons, 1994)</th>
<th>Students’ perspectives (Gibbons, 2003)</th>
</tr>
</thead>
<tbody>
<tr>
<td>To let the students acquire knowledge from the school teacher</td>
<td>To develop students’ skill</td>
</tr>
<tr>
<td>To let the students learn how to teach oneself</td>
<td>To achieve best performance with additional challenges</td>
</tr>
<tr>
<td>To let the students learn how to direct their learning individually</td>
<td>To be self-managed</td>
</tr>
<tr>
<td></td>
<td>To be self-motivated and be able to assess the learning outcome on their own</td>
</tr>
</tbody>
</table>

From the two perspectives mentioned in Table 1, the study by Gibbons (1994) emphasized on how to let the students acquire certain skills from the teacher, while his later study (Gibbons, 2003) highlighted the students’ personal development, such as self-management or self-assessment. This could be referred that the weight for the feasibility of SDL in classrooms might be transformed from teachers’ perspectives to the students’, but we believed that there existed certain reasons that both the teacher and students’ perspectives should be taken into consideration. For that reason, the design of this study would enhance the students’ individuality and teacher’s management for students’ learning experience.
Accordingly, based on the idea of SDL by various researchers, a few studies discussed their application for SDL with ICT. For example, Robertson (2011) applied a blog-based system for students to learn in a self-directed way. In addition, followed by the master plan for education in the country, Teo et al. (2010) developed a questionnaire for SDL with ICT readiness by two pilot studies. Tan, Divaharan, Tan, & Cheah (2011) viewed SDL as a natural learning process, in which students’ ownership, teachers’ monitoring and management for students’ learning were discussed at the elementary level, and they provided practical examples or experience for overcoming the adoption on SDL with ICT in education. Consequently, using the applications in SDL with ICT as references, the design of this study would be further discussed in Section 3.

2.2 Self-efficacy and SDL

Since students have to set their goals and strive for self-directness in SDL, students’ self-efficacy, which indicates students’ personal perception to success (Coutinho & Neuman, 2008), was believed as a positive relation with SDL (Kim, 2009). For the studies related to self-efficacy, most of the studies were derived from Bandura’s Social Learning Theory (Bandura, 1997), in which self-efficacy refers to one’s belief and capabilities for the learning outcome. Also, many researchers examined the difference for self-efficacy between high and low achievement students (Shell, Bruning, & Colvin, 1995; Stephenson, Poissant, & Dade, 1999; Stipek, 2002). In their studies, the researchers concluded that different levels of self-efficacy would result in a difference on performance expectations, because students with high self-efficacy would set a higher goal, task persistence, apply effective learning strategies, and time management than the students with low self-efficacy.

Besides, as the emergence of ICT in education show positive outcome for effective learning, studies discussed the self-efficacy with ICT. For example, Teo (2009) examined the technology acceptance by the pre-service teachers on self-efficacy, while the self-efficacy was believed to have a direct effect related to the ICT readiness, which might affect the learning outcome for the learning activity. The other study, Liaw (2008) also provided evidence that the self-efficacy is also a critical element that affects students’ usage with ICT. Therefore, it is important to examine students’ self-efficacy in SDL. With regards to the discussion above, this study will explore the effect of different levels of self-efficacy, in terms of students’ pace and performance.

3. Design

The essence of SDL emphasizes on students’ individualization. To this end, this study designs a SDL environment based on the design principles originally proposed by Knowles (1975). With regard to the discussion in Section 2, this study designs mechanisms for SDL used in classrooms. In figure 1, the three components: students, the teacher, and the system would be described as follows:

![Figure 1. The design framework in this study.](image-url)
3.1 Students: learn in a self-directed way

Students learn in SDL way, in which interactive learning missions (Chen, Liao, Cheng, Yeh, & Chan, 2012) would be provided as the major learning material during the learning activities. Most of the learning missions are designed based on the spiral curriculum, and these learning missions), along with some drill and practice games (Ku, Chen, Wu, Lao, & Chan, 2014), could be formulated by the concept understanding and procedural fluency (Kilpatrick, Swafford, & Findell, 2001). In the first 40-minute class of every week, students would first review their finished learning missions, then they have to decide how many learning missions and the performance should reach during the learning activity. During this learning process, students would be able to review their needs as well as reflect their learning experience, and set an appropriate goal.

3.2 Teacher: become a facilitator or a problem-solver

For the teacher, s/he is required to look into students’ needs, and offer suitable directions for students to solve the problem in learning. For example, since most students could be able to work on their learning missions individually, a few students, especially low-achievers or those are easily distracted, are needed to be “coached” one-by-one. Moreover, according to the information provided by the system, the teacher would be able to support or provide guidance to the students, offer questions scaffolds that help resolve the mathematical problems, review and negotiate students’ learning goal via the learning system, and maximizing the classroom management (such as offering bonus).

3.3 System: play as a learning portfolio for fulfilling the needs

The computer technology would make the SDL adoption in classrooms easier. Compares to the past, teachers or instructors were needed to review students’ learning process one-by-one. This was quite time-consuming and the classroom teacher were barely to evaluate all students’ learning carefully. Therefore, the design of this study takes the advantage of ICT, by offering real-time database inquiry for both students and the teachers. On the one hand, students can retrieve the status of their learning experience, which helps review and reflect their learning. On the other hand, teachers could be support by the system, in which the every students’ learning status (such as accomplished learning missions, goals, and performance) can be accessed via the teachers’ monitor.

4. Method

This section describes how the study is conducted. Students were randomly distributed in every class, and most of the students (> 90%) brought their parent-purchased tablet PC to the school. Every time when the mathematics class starts, students would turn on their table PCs and visit the web-based learning platform on their own. Nonetheless, due to the fact that most students own different self-efficacy in learning, it is needed to understand the different levels of self-efficacy among the students. To this end, since various studies applied the Motivated Strategies Learning Questionnaire (MSLQ, Pintrich, 1990) for understanding students’ difference, this study adopted the self-efficacy for learning & performance scale in MSLQ. However, the original manuscript of MSLQ is used for adults and is in English, this study applied a localized version by Hsin, Lin, Yeh (2005), and the Cronbach alpha is 0.72.

Before the activity starts, the localized MSLQ was distributed to the students. Thirty-one effective samples were returned from two Grade 2 classes in a public school (See Table 2). As a result, to classify the effects of different levels of self-efficacy, students were categorized into two groups: high and low level of self-efficacy. Students with the above average score would be considered as high self-efficacy group, while the students with below average would be assigned to the low self-efficacy group. In addition, in order to understand the effect on self-efficacy, the performance and the pace were collected in both the pre-test and the post-test.
Table 2: Gibbons’ two studies for facilitating SDL in classrooms.

<table>
<thead>
<tr>
<th>Group of students</th>
<th>Class A</th>
<th>Class B</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>HiSE group</td>
<td>10</td>
<td>6</td>
<td>16</td>
</tr>
<tr>
<td>LoSE group</td>
<td>7</td>
<td>8</td>
<td>15</td>
</tr>
<tr>
<td>Total</td>
<td>17</td>
<td>14</td>
<td>31</td>
</tr>
</tbody>
</table>

5. Results

In order to understand how students’ self-efficacy affects students’ performance and pace, this section describes preliminary analysis for both the pace and performance with independent t-test. To this end, a group of 31 students were divided into two groups based on the quantitative result by the MSLQ questionnaire (see Section 3), i.e. high and low self-efficacy groups (HiSE and LoSE), separately. For the students who own an above-average result, will be assigned to HiSE group (n=16), while the students who own a below-average result, will be the LoSE group (n=15). In this study, the result demonstrates that HiSE students do not have a significant difference with the LoSE students on both performance (t = -1.105, p>.05) and pace (t = .984, p>.05). It implies that the students’ differences on different levels of self-efficacy would not be a factor that might affect students’ learning performance and pace. The reason to this phenomenon might due to the limited flexibility for individuals in the public classrooms, and consequently no significant difference was found on students’ performance and pace.

Table 3: The comparison of pace between high and low self-efficacy groups with independent t-test.

<table>
<thead>
<tr>
<th>Group of students</th>
<th>Mean (number of missions)</th>
<th>SD</th>
<th>d.f.</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>HiSE</td>
<td>240.13</td>
<td>11.05</td>
<td>29</td>
<td>-1.105</td>
<td>.278</td>
</tr>
<tr>
<td>LoSE</td>
<td>245.67</td>
<td>9.20</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

First, as stated in Table 3, the result for students’ pace shows that the HiSE group has no significant difference with the LoSE group, and the HiSE group (mean: 240.13) is slightly lower than the LoSE group (mean: 245.67). This implies that HiSE students would not be beneficial by the difference on the self-efficacy with LoSE students. It might be due to the fact that some of the students in LoSE group have reached the maximum pace (eight students, the maximum pace: 256), and while most of the students in HiSE group is comparatively slower. Second, for the performance (See Table 4), no significant difference is found between the two groups of students. This implied that students with high self-efficacy would not result a difference with the low self-efficacy students. Nevertheless, we notice that an increase on the gain score for the performance of HiSE students, but a decrease for the performance of LoSE students. It could be explained that although no significant difference is found for the gain score, students with different levels of self-efficacy might lead to a slightly difference for the learning performance.

Besides, the average pace and performance in the samples for high self-efficacy have a slightly lower score than the low self-efficacy students. We believed that it might be due to the ceiling effects or some outliers existed among the students.

Table 4: The comparison of performance between high and low self-efficacy groups with independent t-test.

<table>
<thead>
<tr>
<th>Group of students</th>
<th>Mid-term (Mean)</th>
<th>Final (Mean)</th>
<th>Gain score (Mean)</th>
<th>s.d.</th>
<th>d.f.</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>HiSE</td>
<td>86.50</td>
<td>87.06</td>
<td>0.56</td>
<td>4.163</td>
<td>29</td>
<td>.984</td>
<td>.333</td>
</tr>
<tr>
<td>LoSE</td>
<td>90.73</td>
<td>89.87</td>
<td>-0.81</td>
<td>3.907</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
6. Conclusion

With regard to the emerging importance for SDL in the education, students' learning capabilities on different perspectives (such as individual differences, learning habit) should be considered. Therefore, this study provides a preliminary exploration on the self-efficacy for both the performance and the pace in the SDL environment. Hence, in order to understand the effects between students with different levels of self-efficacy, this study addresses a SDL scenario into the public classrooms in Taiwan. As a result, students with different levels of self-efficacy did not show significant differences on both performance and pace. This indicates that the self-efficacy would not be an index for understanding students’ difference for learning in SDL. However, this study is only a preliminary study and analyze for self-efficacy in SDL. Additional exploration should be carefully addressed. For example, the sample size of this study is too small (HiSE group = 16, LoSE group = 15), and additional samples are needed to be done. Also, Further, this study provides a preliminary evidence and analysis on self-efficacy for SDL with ICT, and could be a reference on the adoption of SDL with ICT in education.

Acknowledgements

The authors would like to thank the National Science Council of the Republic of China, Taiwan, for financial support (MOST 101-2511-S-008 -016 -MY3, MOST 100-2511-S-008 -013 -MY3) and Research Center for Science and Technology for Learning, National Central University, Taiwan.

References

Guthrie, R. V. (2004). Even the rat was white: A historical view of psychology. Pearson Education.
The Effects of Game-based Peer Response on Writing Quality: High-ability vs. Low-ability

Jen-Hang WANG*, Sherry Y. CHEN, Oskar KU & Tak-Wai CHAN

*Department of Computer Science and Information Engineering, National Central University, Taiwan
bResearch Center for Science and Technology for Learning, National Central University, Taiwan
cGraduate Institute of Networking Learning Technology, National Central University, Taiwan

*harry@CL.ncu.edu.tw

Abstract: In this study, we proposed a game-based peer response to enhance student writing. In addition, we also examined how students with different levels of ability react to the game-based peer response. The results revealed that such a game-based peer response could enhance students’ writing quality and narrow the gap between the high-ability students and the low-ability students. Moreover, the effects of the game-based peer response on the improvement of high-ability students’ writing quality was limited and did not reach to a significant level. Conversely, the game-based peer response was able to significantly improve the writing quality of low-ability students, regardless the overall quality or each individual aspect of writing quality.

Keywords: Game-based learning, individual differences, peer response

1. Introduction

Peer response (Elbow, 1973; DiPardo & Freedman, 1988) is also known as peer review, which refers to a collaborative activity, in which learners work together to improve the quality of their works by providing comments for each other. Recently, a lot of researchers investigated the effects of peer response and found its benefits can be classified into four aspects: social, cognitive, affective and linguistic (Rollinson, 2005; Min, 2006). Regarding the social aspect, negotiations used in the process of peer response could enhance students’ communication and collaboration skills (Mendonça & Johnson, 1994). Regarding the cognitive aspect, peer response could not only facilitate students to develop critical and analytical skills for writing (Stoddard & MacArthur, 1993), but also make them have a greater awareness of audience (Lockhart & Ng, 1993). Regarding the affective aspect, peer response could help students reduce apprehension and increase confidence (Leki, 1990) and develop a greater sense of the ownership of the work (Tsui & Ng, 2000). Regarding the linguistic aspect, peer response could help students gain more new ideas and different points of view (Lockhart & Ng, 1993) and improve the quality of their works (Cho & Schunn, 2007).

Due to the aforementioned benefits, peer response has been widely applied to enhance student writing. For example, Sims (2001) attempted to use peer comments to improve children’s expressive writing, and found that peer response enhanced students’ writing fluency. Subsequently, Boscolo and Ascori (2004) attempted to apply peer response to assist children to improve the clarity of their narrative writing. They found that peer response fostered students’ abilities to detect information gaps or inconsistencies in writing. Additionally, Tuzi (2004) also used electronic peer feedback to support the revisions of academic writing. He found that electronic peer feedback assisted students to understand how to structure an essay.

The aforementioned studies demonstrated the effectiveness of peer response. Nevertheless, all of these peer response approaches are still implemented in a traditional education context. As suggested by Tüzün, Yilmaz-Soylu, Karakuş, İnal, & Kızılkaya (2009), students in a traditional educational context may have low motivation. In particular, young students have a limited attention span (Moreno-Ger, Martínez-Ortiz, Sierra, & Fernandez-Manjon, 2008). Therefore, there is a need to use a mechanism that can catch students’ attention and increase their motivation. Among various mechanisms, digital games have transformed the way people learn and make learners have enjoyable
experience (Marsh, 2011). Furthermore, some researchers found that digital games can motivate learners because they raise curiosity and allow learners to be in control of their own learning (Dickey, 2007; Huizenga, Admiraal, Akkerman, & ten Dam, 2009; Papastergiou, 2009). This may be the reason why game-based learning (GBL) emerged in the past ten years.

GBL possesses many positive effects on student learning (Pivec, 2007). In particular, previous research found that GBL could enhance students’ learning motivation. For example, a study by Liu and Chu (2010) compared GBL and non-GBL in ubiquitous context. The results demonstrated that students with ubiquitous games could have better learning motivation than those with a non-gamed method. Thus, game-based learning can be a potential approach to address the problems of students’ low motivation and short attention span. To this end, this study attempts to develop game-based peer response by incorporating GBL into peer response.

However, it is unknown whether such game-based peer response can be appreciated by all learners. This is due to the fact that game-based learning includes a variety of multimedia elements, which may cause cognitive overload. In other words, the game-based peer response delivers feedback via multiple information sources, which may increase students’ cognitive load (Fried, 2008). In particular, learners are diverse so not all of learners have such a capacity to handle cognitive overload, which is usually happened when it is beyond the learners’ capacity (Ang, Zaphiris, & Mahmood, 2007). Accordingly, there is a need to consider whether all learners can cope with such cognitive load. In other words, individual differences become an important issue. Among various individual differences, the diversity in learning abilities greatly affect students’ perceptions (Cheng, Lam, & Chan, 2008), which, in turn, will influence their learning outcomes. Thus, such ability differences may affect how students react to this game-based peer response. Therefore, there is a need to examine the impacts of students’ abilities on their reactions to the proposed game-based peer response.

To this end, the aims of this study are two folded. One is to develop game-based peer response while the other is to examine the effects of the ability differences on students’ reactions to the game-based peer response. To correspond to the aforementioned two aims, two research questions are examined in this study:

(a) What are the effects of the game-based peer response on students’ writing quality?
(b) How do high-ability students and low-ability students react differently to the game-based peer response?

The answers to these two research questions can contribute to develop a deep understanding of how to undertake game-based peer response that can accommodate students’ individual differences. By doing so, both high- and low-ability students can benefit from game-based peer response.

2. Methodology Design

This study was conducted in an elementary classroom. To correspond to the aforementioned research questions, an empirical study was conducted to evaluate the effects of game-based peer response. The details are described in this section, including the implementation of the game-based peer response, participants, a pre-test and a post-test, pedagogical activities, measurement of writing quality, and data analysis.

2.1 The Implementation of Game-based Peer Response

We developed a game-based peer response, where peer response was conducted with various game elements, including game activities and rewards. Regarding game activities, participants need to play as a head of a publisher and manage their own publisher by completing various game activities required by different departments of the publisher (Figure 1), such as editing drafts, giving feedback to their peers’ works, evaluating feedback received, revising their own work, and publishing and promoting their completed works to other peers (Figure 2). The purpose of such a series of game activities was to extend students’ attention span in learning, and, in turn, facilitate themselves to complete target learning tasks. This is because each of the aforementioned game activities is associated with a target task, which has a

429
clear sub-goal to be achieved. In other words, a complex writing process was decomposed into a series of tasks. When students complete a target task, they also achieve its sub-goal and can move to pursue the next sub-goal. Through the accumulation of these sub-goals, the ultimate goal is, accordingly, reached.

Further to a series of game activities, various reward mechanisms were also employed to extend students’ attention span and increase their motivation and participation in learning activities, including virtual currency, leaderboards, and trophies. The virtual currency was used to reward students’ behavior for their level of participation and hard working in each game activity and the earned virtual currency can be used to buy marketing tools for promoting users’ published magazines or to order other peers’ published ones. On the other hand, leaderboards and trophies were used in the responding tasks of the Review department. To ensure the responsibility and motivation of users in participating in responding activities, leaderboards and trophies were introduced as levels of social reputation. More specifically, students would receive various levels of trophies based on their ranking on the leaderboard for their feedback performance. In summary, the intention of developing this game context was to enhance students’ engagement in peer response so that their writing quality can be improved.

2.2 Participants

Twenty-one third-grade elementary students participated in this study. They were aged 8-9 years and were recruited from the same class. In other words, they were taught by the same curriculum and were given the same writing assignments and instruction. Furthermore, they had no experience of peer response prior to taking parting in this study.

2.3 Pre-test and Post-test

To evaluate students’ writing ability, participants needed to take a pre-test and a post-test at the beginning and the end of the experiment, respectively. More specifically, the pre-test was applied to examine students’ prior writing abilities while the post-test was employed to assess their writing ability after taking the game-based peer response. The pre-test and post-test were represented as a composition test, where participants were given a theme-based topic and they were required to complete a narrative composition within an 80-minute period.
2.4 Pedagogical Activities

To help participants know how to act as the providers and recipients of peer response, instructions were given to them based on two guidelines: (a) the interaction between readers and writers proposed by Elbow (1973) and (b) the guidance for peer response proposed by Hansen and Liu (2005). By doing so, the participants could undertake peer response with proper attitudes and procedures. Moreover, they were introduced how to complete writing and responding tasks with the game-based peer response.

Then, all participants were evenly re-allocated into small peer response groups of four or five students. Furthermore, participants conducted a series of activities: (a) to receive writing instruction from their teacher, (b) to make drafts individually with a tablet laptop, (c) to receive feedback instruction from their teacher, (d) to read group-mates’ drafts and give feedback, (e) to evaluate how useful the feedback received, (f) to revise their own drafts based on the feedback from their peers, (g) to collect completed works and publish them as a digital publication, (h) to make promotion for their published works to have opportunities to present their works to more audience. During this process, students would obtain various rewards, depending on their performance in the aforementioned target tasks.

2.5 Measurement of Writing Quality

An assessment mechanism proposed by Yang, Ko, and Chung (2005) was adopted to assess students’ writing quality because it was designed for elementary students, and then appropriate for our participants. This assessment covers five items: (1) elegant words, (2) clear paragraph, (3) coherence, (4) title consistency, and (5) new & original ideas. A five-point rating scale was used for each item. Thus, the total score for a composition was between the lowest score (5 points) and the highest score (25 points). Two raters were recruited to independently evaluate the participants’ writing quality so each student’s final score was defined based on the mean of scores by the raters, of which the inter-rater reliability was found to be Kappa = 0.728 (p < .001). In other words, a substantial level for the measure of agreement between the raters was reached.

2.6 Data Analysis

In this study, we investigated how high-ability students and low-ability students react differently to the game-based peer response. Therefore, students were classified into the high-ability and low-ability students based on the mean scores of the pre-test. Then, an Independent Samples t-test, which is suitable to test “the difference between the means of two independent groups” (Howell, 2007), was used to examine differences between the high-ability students and the low-ability students for the pre-test scores and the post-test scores. On the other hand, Paired Samples t-test, which is appropriate to test the difference between the means of paired samples (Howell, 2007), was employed to inspect differences between the pre-test scores and the post-test scores for the high-ability students and the low-ability students. These aforementioned analyses were undertaken by using SPSS for Windows (version 16.0). The level of significance was set at p < .05 for all comparisons.

3. Results and Discussion

This section is divided into three subsections. The first subsection is to present the writing quality of the high ability students and the low-ability students in the pre-test and the post-test. The second subsection is to describe how the high-ability students and the low-ability students performed differently before and after interacting with the game-based peer response. Then, the third subsection is to present a discussion of why the high-ability students and low-ability students reacted differently to the game-based peer response.

3.1 High-ability students vs. Low-ability students
3.1.1 Pre-test Scores

Regarding the pre-test scores, significant differences existed between the scores from the high-ability students and those from the low-ability students. More specifically, the former significantly outperformed the latter, not only in the aspect of overall quality, but also in the aspects of elegant words, clear paragraph, coherence, and new & original ideas (Table 1). These findings suggested that the high-ability students had a better level of prior writing ability than the low-ability students. Accordingly, the former were more capable to use appropriate words and phrases and organize their paragraph structures, and also were better able to express thoughts in a distinctive ways.

On the other hand, the high-ability students did not perform significantly differently from the low-ability students, in the aspect of title consistence \((p > .05)\). This finding implied that both of them had a similar level of prior ability in this aspect.

Table 1: Writing quality of the pre-test (high-ability vs. low-ability).

<table>
<thead>
<tr>
<th></th>
<th>HA (n = 11) M (SD)</th>
<th>LA (n = 10) M (SD)</th>
<th>Independent samples t-test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall</td>
<td>15.27 (1.90)</td>
<td>11.50 (.85)</td>
<td>5.96***</td>
</tr>
<tr>
<td>Elegant words</td>
<td>2.73 (.79)</td>
<td>1.8 (.42)</td>
<td>3.41**</td>
</tr>
<tr>
<td>Clear paragraph</td>
<td>3.09 (.70)</td>
<td>2.20 (.63)</td>
<td>3.05**</td>
</tr>
<tr>
<td>Coherence</td>
<td>2.82 (.87)</td>
<td>2.10 (.32)</td>
<td>2.55*</td>
</tr>
<tr>
<td>Title consistence</td>
<td>4.00 (.00)</td>
<td>3.90 (.57)</td>
<td>.56</td>
</tr>
<tr>
<td>New &amp; original ideas</td>
<td>2.64 (.81)</td>
<td>1.50 (.53)</td>
<td>3.77**</td>
</tr>
</tbody>
</table>

\( * p < .05, ** p < .01, *** p < .001 \)

3.1.2 Post-test Scores

Regarding the post-test scores, no significant differences \((p > .05)\) existed between the scores from the high-ability students and those from the low-ability students. This implied that both groups had a similar level of posterior writing ability, regardless the overall quality or the other aspects aforementioned after they undertook the game-based peer response (Table 2). In other words, the gap between these two groups was minimized.

However, it is still unclear why the aforementioned gap between the high-ability students and the low-ability students has been narrowed or who can benefit from this game-based peer response. Therefore, it is needed to further explore how the high-ability students and the low-ability students react differently to this game-based peer response. To address this issue, we conducted a comparison between students’ pre-test scores and post-test scores, of which the results are presented in the subsection below.

Table 2: Writing quality of the post-test (high-ability vs. low-ability).

<table>
<thead>
<tr>
<th></th>
<th>HA (n = 11) M (SD)</th>
<th>LA (n = 10) M (SD)</th>
<th>Independent samples t-test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall</td>
<td>16.18 (2.23)</td>
<td>16.80 (2.49)</td>
<td>-.60</td>
</tr>
<tr>
<td>Elegant words</td>
<td>2.73 (.47)</td>
<td>2.70 (.48)</td>
<td>.13</td>
</tr>
<tr>
<td>Clear paragraph</td>
<td>2.91 (.83)</td>
<td>3.40 (.84)</td>
<td>-1.34</td>
</tr>
<tr>
<td>Coherence</td>
<td>3.18 (.87)</td>
<td>3.10 (.74)</td>
<td>.23</td>
</tr>
<tr>
<td>Title consistence</td>
<td>4.55 (.69)</td>
<td>4.60 (.52)</td>
<td>-.20</td>
</tr>
<tr>
<td>New &amp; original ideas</td>
<td>2.82 (.75)</td>
<td>3.00 (.67)</td>
<td>-.58</td>
</tr>
</tbody>
</table>

3.2 Pre-test vs. Post-test

3.2.1 High-ability students

Regarding how high-ability students react to the game-based peer response, small differences were found between their post-test scores and pre-test scores in the aspects of writing quality (Table 3).
However, such differences did not reach a statistically significant level (\( p > .05 \)), apart from title consistence (\( p < .05 \)). In other words, the writing quality of high-ability students was not greatly improved. These findings implied that the effect of the game-based peer response on high-ability students’ writing ability was not obvious enough to improve their writing quality.

Table 3: Writing quality of the pre-test and the post-test (high-ability students).

<table>
<thead>
<tr>
<th></th>
<th>Post-test</th>
<th>Pre-test</th>
<th>Paired samples t-test</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M (SD)</td>
<td>M (SD)</td>
<td>( t )</td>
</tr>
<tr>
<td>Overall</td>
<td>16.18 (.23)</td>
<td>15.27 (.90)</td>
<td>1.61</td>
</tr>
<tr>
<td>Elegant words</td>
<td>2.73 (.47)</td>
<td>2.73 (.79)</td>
<td>.00</td>
</tr>
<tr>
<td>Clear paragraph</td>
<td>2.91 (.83)</td>
<td>3.09 (.70)</td>
<td>-.61</td>
</tr>
<tr>
<td>Coherence</td>
<td>3.18 (.87)</td>
<td>2.82 (.87)</td>
<td>1.00</td>
</tr>
<tr>
<td>Title consistence</td>
<td>4.55 (.69)</td>
<td>4.00 (.00)</td>
<td>2.63*</td>
</tr>
<tr>
<td>New &amp; original ideas</td>
<td>2.82 (.75)</td>
<td>2.64 (.81)</td>
<td>.80</td>
</tr>
</tbody>
</table>

\* \( p < .05 \)

3.2.2 Low-ability students

Regarding how low-ability students react to the game-based peer response, significant differences were found between their post-test scores and pre-test scores, not only in the aspect of overall quality but also in the aspects of elegant words, clear paragraph, coherence, title consistence, and new & original ideas (Table 4). In other words, the game-based peer response is beneficial for the low-ability students in all aspects of writing quality. These findings suggested that the game-based peer response could significantly help low-ability students improve their writing quality.

Table 4: Writing quality of the pre-test and the post-test (low-ability students).

<table>
<thead>
<tr>
<th></th>
<th>Post-test</th>
<th>Pre-test</th>
<th>Paired samples t-test</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M (SD)</td>
<td>M (SD)</td>
<td>( t )</td>
</tr>
<tr>
<td>Overall</td>
<td>16.80 (.24)</td>
<td>11.50 (.85)</td>
<td>5.62***</td>
</tr>
<tr>
<td>Elegant words</td>
<td>2.70 (.48)</td>
<td>1.8 (.42)</td>
<td>5.01**</td>
</tr>
<tr>
<td>Clear paragraph</td>
<td>3.40 (.84)</td>
<td>2.20 (.63)</td>
<td>3.67**</td>
</tr>
<tr>
<td>Coherence</td>
<td>3.10 (.74)</td>
<td>2.10 (.32)</td>
<td>3.35**</td>
</tr>
<tr>
<td>Title consistence</td>
<td>4.60 (.52)</td>
<td>3.90 (.57)</td>
<td>3.28*</td>
</tr>
<tr>
<td>New &amp; original ideas</td>
<td>3.00 (.67)</td>
<td>1.50 (.53)</td>
<td>4.39**</td>
</tr>
</tbody>
</table>

\* \( p < .05 \), ** \( p < .01 \), *** \( p < .001 \)

3.3 Discussion

The results of independent samples t-tests presented in the previous subsections suggested that the game-based peer response could narrow the gap between the high-ability students and the low-ability students. More specifically, the change of this gap was from a significant level to a non-significant level. Furthermore, the aforementioned results of paired samples t-tests revealed that the effect of the game-based peer response on the improvement of high-ability students’ writing quality was limited and did not reach a significant level. Conversely, the game-based peer response was able to significantly improve the writing quality of low-ability students, regardless the overall quality or each individual aspect of writing quality. In brief, the low-ability students, but not the high-ability students, benefited from the game-based peer response. Due to such benefits, the former could demonstrate a similar level of writing quality as the latter at the post-test finally.

The fact that the high-ability students and the low-ability students reacted differently to the game-based peer response may be caused by the levels of ability that students possess. In general, high-ability students can obtain the sense of achievement from their works so they enjoy undertaking peer response from intrinsic motivation. Therefore, the impacts of the game context on the enhancement of high-ability students’ motivation in peer response were limited. That is why there are no significant
differences between the high-ability students’ pre-test scores and post-test scores in most aspects of writing quality.

On the other hand, the low-ability students with the game-based peer response showed significantly better performance in all the aspects of writing quality, including elegant words, clear paragraph, coherence, title consistence, and new & original ideas. It may not be easy for the low-ability students to get the sense of achievement from their works so there is a need to drive them by stimulating their extrinsic motivation. This may be the reason why the low-ability students had significant improvement in writing quality after interacting with the game-based peer response. In other words, the game-based peer response played as a mechanism that can stimulate their extrinsic motivation.

More specifically, such a playful mechanism used various game elements to motivate students. For example, the game-based peer response used points to help students get rewards when they accomplished peer response activities. The other game element is the leaderboards, which show how useful students’ comments to their classmates are. By doing so, students’ efforts in providing helpful commentary and criticism could be well recognized. In other words, these game elements could not only enhance low-ability students’ motivation, but also let them have a stronger sense of achievement. Accordingly, the low-ability students demonstrated better progress with the game-based peer response in all the aspects of writing quality.

4. Conclusions

We proposed a game-based peer response to enhance student writing and investigated how students react to the game-based peer response. In addition, the levels of ability that students possess were also considered in this investigation. Therefore, two research questions were examined in this study. Regarding the first research question, i.e., what are the effects of the game-based peer response on students’ writing quality, the results revealed that the game-based peer response could help students enhance their writing quality. Moreover, such a game-based peer response could reduce the gap between the high-ability students and the low-ability students from a significant level to a non-significant level.

Regarding the second research question, i.e., how high-ability students and low-ability students react differently to the game-based peer response, the results suggested that the game-based peer response was significantly beneficial to the low-ability students, but not to the high-ability students, in terms of writing quality. These aforementioned results are interesting but this study is conducted with a small-scale sample. Therefore, future work needs to be undertaken with a large-scale sample to provide more evidence.

Acknowledgements

The authors would like to thank the National Science Council of the Republic of China, Taiwan, for financial support (MOST 101-2511-S-008 -016 -MY3), and Research Center for Science and Technology for Learning, National Central University, Taiwan.

References


The Effects of Mini-Games on Students’ Confidence and Performance in Mental Calculation

Oskar KU*, Denise H. WUb, Andrew C.-C. LAOc, Jen-Hang WANGd & Tak-Wai CHAN*

aResearch Center for Science and Technology for Learning, National Central University, Taiwan
bInstitute of Cognitive Neuroscience, National Central University, Taiwan
cGraduate Institute of Network Learning Technology, National Central University, Taiwan
dDepartment of Computer Science & Information Engineering, National Central University, Taiwan

*oskar@cl.ncu.edu.tw

Abstract: Low confidence toward mathematics is one of the critical issues that diminish students’ will to learn mathematics. Several studies indicated that game-based learning (GBL) might be a potential approach to address this issue. In addition, mental calculation, a fundamental mathematical skill, was considered to be a potential skill that may build students’ confidence toward mathematics. Thus, this study attempted to promote students’ mathematical confidence by incorporating mini-games, a sub-genre of games, into mental calculation learning. A preliminary study was conducted to investigate how students reacted to learn mental calculation in a mini-game environment. The results of the preliminary study suggested that the mini-game-based approach was popular to students. However, it was observed that students’ with different levels of academic ability might benefit diversely from the mini-game environment. Thus, an experiment was conducted to investigate how students with different levels of academic ability react to learn mental calculation with mini-games, in terms of two important aspects of learning, i.e., performance and confidence. A control group participated in a paper-based learning approach to provide a baseline of comparison. The results indicated that The mini-games helped both high- and low-ability students gain significant improvement on their mathematical confidence. In addition, the low-ability students learned with mini-games gained more improvement on mental calculation than their paper-based peers did.

Keywords: mini-games, confidence, math learning, mental calculation, learning performance

1. Introduction

Mathematics is an important and fundamental skill taught in schools. However, it is also a difficult subject in students’ mind (Stodolsky, Salk, & Glaessner, 1991). Such a negative perception may result from students’ low confidence toward mathematics, which, in turn, may diminish students’ will to learn mathematics (Brown, Brown, & Bibby, 2008). The importance of self-confidence can be seen in many aspects (Linnenbrink & Pintrich, 2003; Maclellan, 2014). For example, a person’s self-confidence toward a subject can predict not only how much effort he/she will pay to learn the subject but also his/her expectation of learning outcomes (Schunk, 1990). In addition, high self-confidence may lead students to engage in a learning task actively (Gushue, Scanlan, Pantzer, & Clarke, 2006) and attain better learning outcome (Kleitman, Stankov, Allwood, Young, & Mak, 2013). Thus, helping students to build their confidence toward mathematics is an urgent issue.

Digital games, as an alternative learning approach, may be a possible solution to the problem of students’ low confidence toward mathematics. There have been a number of studies reporting the positive effects of GBL (Pivec, 2007; Chang, Wu, Weng, & Sung, 2012), such as enhancing students’ learning motivation (Klawe, 1998; Nussbaum, 2007) and improving students’ learning performance (Girard, Ecalle, & Magnan, 2012). Importantly, several studies indicated that GBL might enhance students’ confidence (Cunningham, 1994; Radford, 2000). On the other hand, mental calculation, a relatively basic mathematical skill, may also help students shape their confidence toward learning mathematics (Rubenstein, 2001).
To this end, this study aims at investigating whether embedding learning content in digital games can enhance students’ confidence and performance in mental calculation, especially for low confident students. Among various genres of digital games, mini-games are chosen as the learning environments in this research due to their simplicity. Regarding learning material, mental calculation was chosen as the learning content for its importance to mathematics. In addition, how students with different levels of ability react to the mini-games-based learning approach was also examined, because the issue of individual difference is more and more important to the design of digital learning. Thus, the research questions of this study are: (1) Can embedding mental calculation learning in mini-games enhances students’ confidence and learning performance toward mental calculation and their future math learning? (2) Do students with different levels of academic ability react similarly to mini-game-based learning?

2. Related Work

2.1 Mathematical Confidence

Confidence influences students’ performance (Al-Hebaish, 2012) and effort (Bandura, 1982; Tschannen-Moran, Woolfolk Hoy, & Hoy, 1998). For example, confidence was also found to affect one’s math performance (Stankov, Lee, Luo, & Hogan, 2012). Besides, it also affects students’ will to enroll mathematics courses (Metie, Frank, & Croft, 2007).

On the other hand, the recent reports of the Trends in International Mathematics and Science Study (TIMSS) indicated that there were a large proportion of Asian students possessing low confidence toward learning mathematics (Mullis, Martin, Gonzalez, & Chrostowski, 2004; Mullis, Martin, & Foy, 2008). Such a phenomenon was getting more and more serious with the students’ age. This is a critical issue that needs to be noticed because low confidence may make students feel mathematics difficult (Brown, Brown, & Bibby, 2008), which, in turn, may make students avoid facing mathematics. Thus, there is a need to help students build up their mathematical confidence.

2.2 Game-based Learning

Past studies on GBL revealed many positive effects brought by applying digital games to learning environments. For example, GBL enhanced students’ learning motivation (Dickey, 2007) and improved students’ engagement in learning (Huizenga, Admiraal, Akkerman, & ten Dam, 2009). In addition, digital games were found to have the potential to enhance students’ confidence (Cunningham, 1994). For example, Straker et al. (2011) investigated whether playing virtual reality (VR) electronic games helped children with developmental coordination disorder (DCD) gain motor confidence. Their results indicated that playing VR electronic games enhanced DCD children’s confidence in performing motor skills. Such research demonstrated that digital games might have the potential to help students build their confidence toward the course that they are learning. In this vein, digital games may be a possible solution to address the issue of students’ low confidence toward learning mathematics. Therefore, embedding math learning into digital games may be a possible solution to enhance students’ self-confidence toward learning mathematics.

2.2.1 Mini-games

Among various game genres, a mini-game (causal game) is a kind of relatively simple games designed for players who do not want to spend much time and effort on playing games. The main purpose of mini-games is providing people a period of relaxation time between two formal tasks. In contrast with large-scale games, mini-games are simpler and easier to be operate. Even so, they still possess the characteristics of a game. In most computer games, players have to devote significant attention on playing games. However, in mini-games, players can keep more concentration on learning tasks rather than playing games; games are just assistants that engage students in learning tasks. On the other hand, because mental calculation requires a high level of concentration for students to produce an answer, there is a need to reduce students’ cognitive load. To this end, mini-game is adopted as the design of the
games with exchangeable learning materials for this study. More specifically, mini-games were adopted majorly due to their simplicity, which might cause less cognitive load for learners than complex games did. Additionally, mini-games are content-independent, which makes it possible for the mini-games to be associated with different learning materials.

3. Research Design

Two studies were conducted to investigate the effects of mini-games on students’ confidence and performance in mental calculation. In addition, a framework is presented to illustrate the relationship between the two studies.

3.1 Preliminary Study

Since not all GBL produced positive results in past studies, there is a need to examine whether such a pedagogy is beneficial for building students’ confidence toward mental calculation and enhancing their performance. Thus, a preliminary study was conducted to examine whether learning in mini-game environments brings positive effects on students’ confidence and performance in mental calculation. The purposes of the preliminary study included the following aspects:
1) Examining the effects of mini-game-based mental calculation learning. If students’ attitude and performance on mental calculation were improved due to the intervention of mini-game-based learning (mGBL), a study with experimental design would be conducted to investigate how the mGBL approach builds students’ confidence and meanwhile enhances their performance subsequently.
2) Collecting students’ learning behavior. Students’ reactions to the mini-game-based mental calculation learning will be analyzed to serve as the design principles for improving the game design in the subsequent study.

3.2 Main Study

Based on the results of the preliminary study, a main study was designed and conducted to verify whether the game-based approach significantly build students’ confidence toward learning mental calculation and improve their mental calculation performance. Importantly, the main study further investigates whether the mini-game-based approach brings different learning effects on students with different levels of learning ability.

![Figure 1](image-url)  
*Figure 1. The relationship between Study 1 & Study 2.*

3.3 Research Framework
Figure 1 illustrates the relationship between the preliminary and the main studies. The preliminary study explored the possible effects of game-based mental calculation learning while the main study further investigated how the effects were produce and whether the effects acted equally on students with different levels of ability.

4. Study One: Preliminary Study

4.1 Participants

The participants of the preliminary study were fourth grade students from an elementary school in northern Taiwan. Students in the school were normally distributed to the classes of the school based on their academic performance. Thus, a class with 28 students was randomly selected to participate in the preliminary study.

4.2 Instruments

4.2.1 Mini-games

Four mini-games (Fig. 2) were developed as the learning environments for students to learn the mental calculation skills.

![Game A](image1)  ![Game B](image2)  ![Game C](image3)  ![Game D](image4)

Figure 2. Screenshots of mini-games

The four mini-games were designed as two multiple choice questions games (A & B) and fill-in questions (C & D). The detail descriptions of the mini-games are shown in Table 1.

Table 1: The descriptions of the mini-games.

<table>
<thead>
<tr>
<th>Game</th>
<th>Name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Space Traveler</td>
<td>Problems are shown at the top of game screen. Students have to shoot the asteroid with correct answer. Otherwise, their spacecraft will crash.</td>
</tr>
<tr>
<td>B</td>
<td>Forest Protector</td>
<td>Students have to shoot the invader who holds the correct answer to the question displayed at the top of game screen.</td>
</tr>
<tr>
<td>C</td>
<td>Light City</td>
<td>Students have to enter the correct answer to the question shown at the bottom of game screen to prevent the bomb cart from exploding.</td>
</tr>
<tr>
<td>D</td>
<td>Panda Math</td>
<td>Students have to help the panda in the screen answer the question thrown by the enemy before the question block hit the panda.</td>
</tr>
</tbody>
</table>

4.2.2 Materials

In order to prevent the interference from regular math courses, this study adopted mental calculation, which was not taught in regular math courses, as students’ learning materials.
There are thousands of mental calculation strategies to simplify computational problems. In addition, students can also invent their own strategies. In this study, several patterns of mental calculation skills were selected according to students’ prior knowledge. The goal of this study is helping students gain more “number sense”, so that they might get more sense of analyzing numbers and consequently formulate an algorithm to simplify the problem they met.

4.3 Procedures

In order to investigate students’ reactions to learning mental calculation within the mini-games, a set of instruction was delivered to the participants twice a week. The set of instruction lasted for nine weeks. The Procedures of each session are depicted in Figure 3. In the beginning of each session, the teacher conducted a mini lecture to introduce a mental calculation skill for five minutes. Then, the teacher used five minutes to ensure whether students understood the content of the mini lecture. Finally, students practice in mini-games for 20 minutes.

Figure 3. The procedures of a session

4.4 Results and Discussion

According to students’ math scores of their school exam, students were divided into high- and low-ability groups (Table 2) to investigate whether they react differently to the intervention of mini-game-based learning. In order to investigate the impact of the mini-games in this study, the accuracy of students with different math abilities and the response time (RT) of students with different math abilities were analyzed.

Table 2: The numbers of the participants in the low and high ability groups.

<table>
<thead>
<tr>
<th>High ability</th>
<th>Low ability</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>13</td>
<td>15</td>
<td>28</td>
</tr>
</tbody>
</table>

4.4.1 Accuracy

The result of the Mann-Whitney test revealed a significant difference of the gain score of accuracy between the high- and the low-ability students (U = 52.000, p = .036). Such a result may imply that the low-ability students gained more improvement than high-ability students in terms accuracy (Figure 4). Furthermore, according to students’ mathematical achievements, the learning profile of the first three, the middle three, and the last three students were chosen to be analyzed. Figure 5 illustrates the average accuracy of the nine students. High math achievement students gained little improvement in
accuracy; they performed very well from the beginning of the learning sessions. The middle and left-behind group students gained quite steady improvement on the accuracy of answering questions.

4.4.2 Response Time

Students’ change of response time is shown in Figure 6. The result of the Mann-Whitney test indicated that no significant difference of the gain score of response time was found between the high- and the low-ability students (U = 76.000, p = .322).

![Figure 6. The change of response time](image)

However, when looking into the response time of the first three, the middle three, and the last three students, we can find that students showed quite different patterns of the change of response time. Figure 7 depicts students’ RT. In the high achievement group, all the three students gained improvement in RT steadily. In the middle and the left-behind groups, the improvement seemed unstable. The differences among all sessions were large. Sometimes they may performed very well in a session but very poor in the next time. Maybe it was due to the fact that the learned skills in left-behind student were not consolidated enough; thus, those student performed quite unstable in their response time on answering questions.

Although several positive effects were obtained in this study, an important issue emerged with the conduction of this preliminary study should be further investigated: students with various abilities demonstrated different levels of improvement in mental calculation skills. A further study needs to be done to investigate whether such a result was caused by GBL.

5. Study Two: Main Study

The main study was conduct to investigate the ability issue raised in the preliminary study. More specifically, since individual difference is a key factor when designing individual learning environments, this study further investigates how students with different levels of academic ability react to mGBL on mental calculation in terms of two important aspects of learning, i.e., performance and confidence.

5.1 Experimental Design

A quasi-experiment was conducted to examine the effects of mGBL. Participants were fourth–grade elementary school students from Northern Taiwan. The participants were 59 students (N=59), aged 10-11 years old. Two classes of students were randomly selected and then assigned as an Experimental Group (EG, 14 males and 17 females) and a Control Group (CG, 12 males and 16 females). The experimental group received computer game-based learning, while the control group received paper-based learning.

In addition, students in both groups were further divided into two subgroups—high achievement and low achievement—to investigate the effects of mGBL on high and low achievement students.
5.2 Instruments

The learning materials used in the main study were the same as those in the preliminary study. Students in the EG learned in mini-game environments while CG students learned with paper handouts and worksheets.

5.2.1 Mini-games (EG)

Two mini-games were adopted as the learning environments for EG students to interact with learning materials. The two mini-games were implemented as a board game and a sports game (Fig. 4). The rules of both games are simple so that players can get started with the games without difficulty. Students learned and practiced in the mini-game environments. Students can replay the electronic learning materials at any time.

![Figure 4. The board game and the sports game](image)

5.2.2 Paper Handouts and Worksheets (CG)

Every student in the CG received a copy of learning materials printed on A4 paper. After receiving a short instruction from their teacher, students were asked to practice the newly learned mental calculation strategies with their worksheets. They could review the learning material whenever they have problems during practice. They could also ask their teacher to explain the mental calculation strategy individually. The teacher can pause students’ practice and re-explain the strategy to the whole class if he found too many students encountered the same problem.

5.3 Procedures

The procedure of this study is presented in Figure 5. The pretest was administrated a week before the experimental intervention. The posttest was administrated a week after the last session of the experimental intervention. During the experiment, nine learning sessions were conducted for five consecutive weeks; each session lasted for 25 minutes. In each session, the teachers in both groups taught students the key features of a new strategy or reviewed previous strategies for five minutes. Students then preceded individual learning for 20 minutes.

![Figure 5. The procedures of a session](image)

5.4 Results and Discussion
In order to examine whether students with different levels of ability benefited equally from the mini-game-based approach, participants were divided into four sub-groups in terms of their academic abilities (Table 3).

Table 3: The numbers of students grouped by their math ability.

<table>
<thead>
<tr>
<th>Ability</th>
<th>EG</th>
<th>CG</th>
</tr>
</thead>
<tbody>
<tr>
<td>High ability</td>
<td>14</td>
<td>12</td>
</tr>
<tr>
<td>Low ability</td>
<td>12</td>
<td>13</td>
</tr>
</tbody>
</table>

5.4.1 Confidence

As shown in Table 4, low-ability CG students demonstrated a negative trend of confidence change, while the other three sub-groups demonstrated a positive change of confidence. In other words, low-ability CG students’ confidence dropped after experiment.

Table 4: The mean scores of confidence toward mathematics grouped by ability.

<table>
<thead>
<tr>
<th>Ability</th>
<th>Pretest</th>
<th>Posttest</th>
<th>Z</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
<td>SD</td>
</tr>
<tr>
<td>CG</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low</td>
<td>3.19</td>
<td>0.57</td>
<td>2.75</td>
<td>0.76</td>
</tr>
<tr>
<td>High</td>
<td>3.55</td>
<td>0.67</td>
<td>3.69</td>
<td>0.88</td>
</tr>
<tr>
<td>EG</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low</td>
<td>3.14</td>
<td>0.48</td>
<td>3.45</td>
<td>0.56</td>
</tr>
<tr>
<td>High</td>
<td>3.57</td>
<td>0.54</td>
<td>3.86</td>
<td>0.81</td>
</tr>
</tbody>
</table>

Regarding the mini-game-based (EG) condition, the Wilcoxon’s matched-pairs signed-ranks tests revealed that both high- (Z = 2.165, p = .030) and low-ability (Z = 2.156, p = .031) students gained significant improvement on confidence. In addition the Mann-Whitney test indicated no significant difference of the gain score of confidence between the two sub-groups (U = 81.000, p = .876). Conversely, both high- and low-ability CG groups did not demonstrate significant improvement on their confidence.

For low-ability students, such a result may be caused by the positive feedback (scores, reward, and correct signs) provided by mini-games. Frequently receiving a message about completing a learning task successfully may encourage low-ability students to build their confidence toward similar tasks in the future (Pajares, 2006).

As for the high-ability EG students, there may be another source for their improvement of confidence—challenge. High-ability students may tend to expect challenge (Jones & Spooner, 2006). When completing challenging tasks, students may obtain a sense of achievement (Dickey, 2007), which, in turn, may enhance their confidence (Hammond, 2004). The mini-games adjust the game challenge dynamically; thus, high ability students can always receive adequate challenge during gameplay.

5.4.2 Mental Calculation Performance

Table 5 presents students’ mental calculation performance. The Wilcoxon’s matched-pairs signed-ranks tests indicated that students in each sub-group gained significant improvement of their mental calculation performance.

Table 5: The mean scores of confidence toward mathematics grouped by ability.

<table>
<thead>
<tr>
<th>Ability</th>
<th>Pretest</th>
<th>Posttest</th>
<th>Z</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
<td>SD</td>
</tr>
<tr>
<td>CG</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low</td>
<td>10.85</td>
<td>2.94</td>
<td>36.08</td>
<td>20.89</td>
</tr>
<tr>
<td>High</td>
<td>33.58</td>
<td>25.29</td>
<td>76.08</td>
<td>25.49</td>
</tr>
<tr>
<td>EG</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High</td>
<td>24.64</td>
<td>22.20</td>
<td>75.21</td>
<td>22.08</td>
</tr>
</tbody>
</table>
For the CG students, as expected, the result of the Mann-Whitney test demonstrated a significant difference of the gain score of mental calculation performance between the high-ability and the low-ability students (U = 40.000, p = .039). In other words, the high-ability students in the CG gained more improvement than their low ability peers in the same setting.

On the other hand, although the high-ability EG students did not perform as well as their high-ability CG peers, they attained a comparable level of performance in the posttest. As for low-ability students, the result of the Mann-Whitney tests demonstrated a significant difference of the gain score between the EG low-ability and the CG low-ability students (U = 38.500, p = .032) while the difference of pretest between the EG low-ability and the CG low-ability students was not found (U = 46.500, p = .085). This result might imply that the low-ability EG students gained more improvement than those in the CG.

The results demonstrated in this section indicated that the mini-game-based learning might provide greater benefit to low-ability students in terms of the learning gain. Such a result might be caused by the immediate feedback for error correction in the mini-games, which helped learner consolidate knowledge (McDaniel, Roediger, & McDermott, 2007).

6. Conclusions

Confidence plays an important role in students’ learning process. However, a lack of confidence toward mathematics is many students’ common problem, which prevents them from pursuing advanced mathematical knowledge. This research introduced mini-games as an approach to motivate students to learn a fundamental mathematical skill, mental calculation and investigated whether such an approach can enhance students’ confidence toward mathematics.

The results from both the preliminary and the main studies indicate that students’ confidence and performance in mental calculation were improved after receiving a mini-game-based mental calculation course. Furthermore, the experimental results obtained from the main study indicate that students with the mini-game-based environment gained significant improvement on confidence toward mathematics; both high and low ability students gained significant improvement on their mathematical confidence. In contrast, students learned with the paper-based setting did not demonstrate similar change on their mathematical confidence. On the other hand, students in all conditions gained significant improvement for mental calculation performance. Importantly, the EG students gained more improvement than their CG peers.

Acknowledgements

The authors would like to thank the National Science Council of the Republic of China, Taiwan, for financially supporting this study under contract numbers NSC 100-2511-S-008-014-MY3 and NSC 100-2511-S-008-013-MY3.

References


The Interface Design of Electronic Journals via Mobile Devices: A Cognitive styles Perspective

Chu-Han Chan & Sherry Y. Chen

Abstract: With the advancement of information technology, combining with electronic journals and mobile devices would produce ubiquitous electronic journals. However, there is a need to consider the usability evaluation because usability is a strong predictor of design issues. To satisfy individual needs, the effects of cognitive styles on usability inspection are investigated in this study. To this end, this study aimed to examine how different cognitive style groups perceive the interface design of an electronic journal. More specifically, Nielsen’s ten heuristics (Hs) were applied to investigate user’ perceptions. The results show that H8 was considered the most important heuristic by all users. The results also demonstrate that Holists who perceive excessive advertising may strongly need previous/next buttons while Serialists who feel this electronic journal provides too many advertising may consider that too much information is presented in the home page. The findings can be applied to support the development of individualized mobile electronic journals.

Keywords: Electronic journals, Mobile devices, Cognitive styles, usability, Nielsen’s heuristics

1. Introduction

Digital learning refers to utilize digital technologies to support student learning (Chan, et. al., 2006). Among a variety of digital technologies which can be applied to implement learning materials, mobile devices particularly offer many advantages, e.g., convenience, flexibility and ubiquitous information access (Jacob & Issac, 2008). Among these advantages, the portability is a major advantage that leads to the other two. Regarding flexibility, portability can facilitate users to access information anytime (Liu and Carlsson, 2010). Regarding ubiquity, portability removes geographic boundaries so users can locate information at any locations (Looney et al., 2004). Due to these advantages, there are on-going interests to use mobile devices to support teaching and learning recently (Morris, 2010; Petrova and Li, 2009). For instance, Wurst, Smarkola and Gaffney (2008) compared ubiquitous mobile learning with a traditional lecture-based course in higher education. The results from their study suggested users with mobile learning showed significantly more satisfaction than those in traditional classrooms. More recently, Cavus and Uzunboylu (2009) used the mobile devices to develop a mobile learning system and they found both users’ attitudes toward the mobile devices and their creativity were improved significantly at the end. In summary, mobile learning does indeed become a mainstream method of education in 21st Century (Peters, 2007).

Further to mobile devices, electronic journals are another useful digital technology widely used in educational settings because they can facilitate to disseminate scientific information (Ollé and Borrego, 2010). By doing so, students can effectively acquire new information to enhance their understandings. In addition to disseminating scientific information, the electronic journals also provide other benefits, including the speed of access and the ability to download, print, and send articles (Tyagi, 2011). Due to the widespread use of electronic journals, research into this issue has mushroomed. In an early period, Bar-Ilan and Fink (2005) conducted a study to examine the use of printed and electronic journals in a science library. The results showed more than 80% of the respondents frequently used and preferred an electronic format. Later on, Prabha (2007) tracked journal subscription and format data for 515 journals in the Association of Research Libraries (ARL) university member libraries. The findings showed journals subscribed in print only decreased to one-third of the journal collections while, concurrently, access to electronic journals increased to one-third of the collections.
The aforementioned studies demonstrated electronic journals played an important role in scholarly communication. Such importance increases the use of electronic journals in various countries. For instance, Kurata et al., (2007) examined the position of electronic journals in scholarly communication based on Japanese researchers’ information behavior. The results showed Japanese researchers used electronic journals for information access as a matter of course. Recently, Bravo and Díez (2011) examined the models of consumption of the academic communities of five Spanish universities. Their study revealed the overall totals for downloads at the universities showed constant growth from 2002 onward. In other words, there was an upward trend in the consumption of scholarly information in electronic formats in the Spanish academic communities.

The aforesaid results demonstrated electronic journals are popular academic tools. In other words, there are an increasing number of users to access electronic journals. On the other hand, great diversities exist among such users, who may have heterogeneous backgrounds, in terms of their knowledge, skills and needs (Chen and Macredie, 2010). Thus, it is necessary to examine relationships between individual differences and the use of electronic journals. Among various individual differences, previous studies mainly focused on examining how users’ subject background affected their information seeking behavior (Talja and Maula, 2003). In addition to subject background, other human factors are also essential, e.g., cognitive styles, which refer to a person’s information processing habits, capturing an individual’s preferred mode of perceiving, thinking, remembering, and problem solving (Messick, 1976). Previous research found cognitive styles are key determinants to affect users’ information seeking (Clewley et al., 2010). Thus, it is necessary to examine how different cognitive style groups react to the use of electronic journals.

Among various dimensions of cognitive styles, Pask’s Holism/Serialism has been received attention recently. Jonassen and Grabowski (2012) describe Holists as preferring to process information in a ‘whole-to-part’ sequence. In contrast, Serialists are described as preferring a ‘part-to-whole’ processing of information. Holists and Serialists have different characteristics. Due to such differences, recent works examined how Holists and Serialists behave differently. For instance, Clewley et al., (2011) found Serialists and Holists have different preferences for their navigational styles. The former prefer to follow a linear pattern by having a suggested route or looking at the subject content step-by-step with back/forward buttons. Conversely, the latter tend to take a non-linear pattern by ‘jumping’ between different levels of subject contents with hypertext links. Furthermore, Chen and Chang (2014) investigated how member grouping affects users’ reactions to mobile collaborative learning from a cognitive style perspective. The results suggest there is a need to provide Serialists with additional help when they use mobile collaborative learning.

In addition to the effect of the cognitive styles, the interface design of the electronic journals is also important because user interface may be thought of as a ‘window’ through which users interact with electronic journals so the design of user interface may affect how users access electronic journals. In other words, the user interface formulates the working environment of electronic journals so it is critical that the working environment is friendly enough to accommodate users’ different preferences. As such, the usability evaluation of electronic journals becomes paramount because it can provide concrete prescriptions for developing electronic journals that are able to align to diverse users’ needs. A number of methods can be used to evaluate usability. Among them, Nielsen’s heuristic approach is most commonly used because it can be used effectively by novices and experts alike and can be performed at any stages of the development lifecycle (Nielsen, 1994a). Nielsen’s Heuristics were first formally described in presentations in the Human–Computer Interaction conference through papers published by Nielson and Molich (1990). Since then, they have refined the heuristics based on a factor analysis of 249 usability problems to derive a revised set of heuristics with maximum explanatory power. Table 1 presents the detail of the revised set of 10 heuristics (H).

These ten heuristics are concise and simple to learn so they are widely applied to evaluate the user interface of a variety of applications. Petrie and Power (2012) assessed the usability of six complex, highly interactive websites based on Nielsen’s heuristics. The results of their study showed there were 935 usability problems found in the evaluation. Recently, Hsieh, Su, Chen and Chen (in press) also used Nielsen’s ten heuristics to assess the usability of a robot-based learning companion. Based on the results of the assessment, they developed three versions of robot-based learning companion. Due to such popularity, the study presented in this paper also assesses the usability of a game-based learning system with Nielsen’s ten heuristics.
Table 1: Nielsen’s ten heuristics (1994b).

<table>
<thead>
<tr>
<th>Heuristics</th>
<th>Explanations</th>
</tr>
</thead>
<tbody>
<tr>
<td>H1: Visibility of system status</td>
<td>The system should always keep user informed about what is going on by providing appropriate feedback within reasonable time</td>
</tr>
<tr>
<td>H2: Match between system and the real world</td>
<td>The system should speak the user’s language, with words, phrases and concepts familiar to the user, rather than system-oriented terms. Follow real-world conventions, making information appear in a natural and logical order</td>
</tr>
<tr>
<td>H3: User control and freedom</td>
<td>Users should be free to develop their own strategies, select and sequence tasks, and undo and redo activities that they have done, rather than having the system do these for them</td>
</tr>
<tr>
<td>H4: Consistency and standards</td>
<td>Users should not have to wonder whether different words, situations, or actions mean the same thing and the system should follow platform conventions.</td>
</tr>
<tr>
<td>H5: Error prevention</td>
<td>Even better than good error messages is a careful design, which prevents a problem from occurring in the first place.</td>
</tr>
<tr>
<td>H6: Recognition rather than recall</td>
<td>Make objects, actions, and options visible. The users should not have to remember information from one part of the dialogue to another. Instructions for use of the system should be visible or easily retrievable whenever appropriate.</td>
</tr>
<tr>
<td>H7: Flexibility and efficiency of use</td>
<td>Allow users to tailor frequent actions. Provide alternative means of access and operation for users who differ from the “average” user (e.g., physical or cognitive ability, culture, language, etc.)</td>
</tr>
<tr>
<td>H8: Aesthetic and minimalist design</td>
<td>Dialogues should not contain information that is irrelevant or rarely needed. Every extra unit of information in a dialogue competes with the relevant units of information and diminishes their relative visibility.</td>
</tr>
<tr>
<td>H9: Help users recognise, diagnose and recover from errors</td>
<td>Error messages should precisely indicate the problem and constructively suggest a solution. They should be expressed in plain language.</td>
</tr>
<tr>
<td>H10: Help and documentation</td>
<td>Even though it is better if the system can be used without documentation, it may be necessary to provide help and documentation. Any such information should be easy to search, focused on the user’s task, list concrete steps to be carried out, and not be too large.</td>
</tr>
</tbody>
</table>

The aforementioned studies demonstrate the usefulness of Nielsen’s heuristic evaluation. However, paucity of studies uses Nielsen’s heuristics to assess the user interface of electronic journals, i.e., the ScienceDirect. In particular, there is a lack of studies to investigate Holists and Serialists’ reactions to electronic journals in the context of mobile devices. To this end, we address this issue. In brief, the aim of this study is to examine how different cognitive style groups perceive the interface design of an electronic journal.

2. Methodology

2.1 Participants

As indicated by Nicholas et al. (2009), the majority users of digital resources were students. Thus, the participants (N=23) were recruited from master students from the Department of Computer Science and Information Engineering at National Central University in Taiwan. In other words, the participants had a similar subject background so that the effects of prior knowledge could be minimized. In addition, a request was issued to students in lectures, and further by email, making clear the nature of the study and their participation. All participants had the basic computer and Internet skills necessary to use the electronic journals.
2.2 *ScienceDirect*

Among various electronic journals, this study adopted the ScienceDirect (Figure 1) to reach the aim described in Section 1. This is because the ScienceDirect covers various topics, such as life sciences, chemistry, and physics. Furthermore, the ScienceDirect also provides multiple search mechanisms: (1) Basic Search, (2) Advanced Search and (3) Expert Search, which differ with respect to the complexity of their interface design and search mechanisms. More specifically, the Expert Search and Advanced Search were considered as an example of complex search design whereas the Basic Search was appreciated by its simplicity. Having such varieties in interface design and search mechanisms provides a wider range of choices, which can help to identify users’ preferences.

![Figure 1. The homepage of the ScienceDirect.](image)

2.3 *Questionnaire*

To investigate how users with different cognitive styles perceived the interface design of the ScienceDirect. A paper-based questionnaire was developed and it included two parts. In the first part, which included 10 three-point Likert-scale questions (“disagree”, “general” and “agree”), users were asked to describe the degree of their satisfaction with the ScienceDirect on the basis of each heuristic. The internal consistency for the overall scale is 0.58 by Cronbach’s alpha, which indicates an adequate satisfaction of the questionnaire. In the second part, which consisted of 30 questions, users were requested to check whether the interface design of the ScienceDirect met the criteria of each heuristic.

2.4 *Experimental Procedures*

To achieve the aim of this study, the procedure included three steps (Figure 2). Initially, all participants were required to fill out their personal information and the SPQ. According to the results of the SPQ, our participants consisted of 12 Holists and 11 Serialists. Subsequently, all of the participants were trained to learn the principles of Nielsen’s heuristics so that all of the participants had the understandings of how to conduct the usability assessment. Then, they were required to interact with the ScienceDirect via tablet PCs. Finally, the participants needed to evaluate the usability of the ScienceDirect based on Neilson’s ten heuristics. Such evaluation was conducted via the questionnaire described in Section.

![Figure 2. The Experimental Procedure.](image)
2.5 Data Analyses

Traditional statistics were applied to conduct data analyses from both macro and micro views in Study Two. The macro view covers two aspects: (a) relationships between the satisfaction of each heuristic and (b) relationships between each criterion in all heuristics. The micro view is obtained by further examining the aforementioned relationships. Spearman’s correlations, which could be used to interpret the strengths of a statistical relationship between two random variables (Stuart et. al, 1991), were applied to find the aforesaid macro view and micro view. Such analyses were undertaken by using Statistical Package for the Social Sciences (SPSS) for Windows (release 18.0). A significance level of p<0.05 was adopted for this study.

3. Results and Discussions

3.1 Overall

The satisfaction of H8 is negativity related to that of H1\(r=-.458, \ p<.05\) and positively related to H3\(r=.492, \ p<.05\) and H6 \(r=.492, \ p<.05\). The results indicated users with high satisfaction with H8 would show low satisfaction with H1 whereas they would show high satisfaction with H3 and H6. In other words, the users’ satisfaction with H8 plays an important role. Thus, this study also conducted detailed analyses for questions related to H8, H1, H3 and H6. As displayed in Table 2, H8 includes three items, i.e., Q24, Q25 and Q26. Q24 is associated with Q18 belonged to H6. Q25 is linked with Q3 and Q8, which are belonged to H1 and H3, respectively. Q26 is connected with Q18 belonged to H6. These findings suggest Q25 is an important issue, which is related to Q3 and Q8. More specifically, too many advertisements may let users feel that it is difficult to identify where the Expert Search is and that there is a need to provide previous/next buttons. The other important issue is Q18, which is related to Q24 and Q26. In other words, presenting too much information in the home page may also make users feel that this electronic journal provides too many functions and too much information. This finding suggests displaying too much information in the home page may cause users’ cognitive overload so they cannot appreciate the value of information and function provided by the electronic journal. In brief, there is a need to pay enough attention to Q18 and Q25, which are essential for the interface design of electronic journals.

Table 2: The variables of Nielsen’s Heuristics (The whole sample).

<table>
<thead>
<tr>
<th></th>
<th>H8</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Excessive functions (Q24)</td>
<td>Excessive advertising (Q25)</td>
<td>Overall Excessive information (Q26)</td>
</tr>
<tr>
<td>H1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Highlighted Keywords (Q1)</td>
<td>.233</td>
<td>.042</td>
<td>.215</td>
</tr>
<tr>
<td>Lack of detailed instruction (Q2)</td>
<td>-.094</td>
<td>-.094</td>
<td>-.210</td>
</tr>
<tr>
<td>Hard to find the location of the Expert Search (Q3)</td>
<td>-.342</td>
<td>-.533**</td>
<td>-.151</td>
</tr>
<tr>
<td>H3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lack of undo/redo functions (Q7)</td>
<td>.279</td>
<td>.058</td>
<td>.128</td>
</tr>
<tr>
<td>Lack of previous/next buttons (Q8)</td>
<td>-.086</td>
<td>.509*</td>
<td>-.066</td>
</tr>
<tr>
<td>Provisions of multiple search (Q9)</td>
<td>-.350</td>
<td>-.163</td>
<td>-.302</td>
</tr>
<tr>
<td>H6</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Too many subject categories (Q17)</td>
<td>.387</td>
<td>.147</td>
<td>.250</td>
</tr>
<tr>
<td>Excessive information in the Home page (Q18)</td>
<td>.707**</td>
<td>.311</td>
<td>.691**</td>
</tr>
<tr>
<td>Clear text icons (Q19)</td>
<td>.042</td>
<td>.042</td>
<td>.032</td>
</tr>
</tbody>
</table>

Keys: * p < .05, ** p < .01

3.2 Cognitive styles

Further to the aforesaid findings for the whole sample, how each cognitive style group reacted to each Nielsen’s heuristic is also analyzed. Holists and Serialists share some similarities but several differences also exist between them.
3.2.1 Similarities

The satisfaction of H8 was positively related to H3 for Holists (r=.622, p<.05). On the other hand, the satisfaction of H8 was positively related to H6 (r=.777, p<.01) and negativity related to H1 (r=-.712, p<.05) for Serialists. These results indicated Holists and Serialists who showed high satisfaction with H8 would show high satisfaction with H3 and H6, respectively but Serialists would also show low satisfaction with H1. In other words, the satisfaction with H8 plays an important role for both Holists and Serialists. Thus, this study also conducted detailed analyses for questions related to H8, H1, H3 and H6. As displayed in Table 3, H8 includes three items, i.e., Q24, Q25 and Q26. Regarding Holists, Q24 and Q25 are associated with Q9 and Q8 belonged to H3. The findings from Holists are similar to those from the whole sample. More specifically, Holists who perceived excessive advertising may strongly need to use previous/next buttons. Additionally, Holists who perceived excessive functions may not need the provision of multiple search. This may be due to the fact excessive advertising and functions increase their cognitive overload already so they do not need multiple search but they need previous/next buttons to facilitate their navigation in hyperspace.

Regarding Serialists, Q25 is related to Q18 belonged to H6 and Q26 is connected with Q1 and Q18 belonged to H1 and H6, respectively. These findings suggest Q26 and Q18 are important issues. Regarding Q26, highlighted Keywords in search results and too much information displayed in the home page may let Serialists feel overwhelmed. Regarding Q18, presenting too much information in the home page may also make Serialists feel this electronic journal provides too many advertising and information. This finding is consistent with the results from 3.1 which claim too much information displayed in the home page may cause users’ cognitive overload. Such a problem may be more serious to Serialists because they only use the options that are relevant to their current tasks (Clewley et al., 2010), which, in turn, they cannot appreciate the value of rich information provided by the electronic journal. In brief, Q18 and Q26 are essential factors for designing the interface of electronic journals for Serialists.

3.2.2 Differences

Regarding Serialists, the satisfaction of H1 was negatively related to H6 (r=-.969, p<.05). Regarding Holists, the satisfaction of H8 was negatively related to H5 (r=-.32, p<.05) and the satisfaction of H1 was positively related to H7 (r=.853, p<.01). In other words, the users’ satisfaction with H1 plays an important role. Thus, this study also conducted detailed analyses for relationships between questions belonged to H1 and those belonged to H6 and H7. However, no significant relationships were found for Serialists. Conversely, some significant relationships were discovered for Holists. As displayed in Table 4, H1 includes three items, among which both Q1 and Q3 are associated with Q22 belonged to H7. In other words, Q22 is an important issue. Regarding Q1, the highlighted keywords in search results may be enough for Holists so that they do not need different types of font size to enhance the visual clue. Regarding Q3, it is difficult to find the location of the Expert Search for Holists so they may need to change the font size to help them find where the Expert Search is located.

Table 3: Findings similar to the whole sample.

<table>
<thead>
<tr>
<th></th>
<th>H8</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Excessive functions (Q24)</td>
</tr>
<tr>
<td>H3</td>
<td>Lack of undo/redo functions (Q7)</td>
</tr>
<tr>
<td></td>
<td>Lack of previous/next buttons (Q8)</td>
</tr>
<tr>
<td></td>
<td>Provisions of multiple search.(Q9)</td>
</tr>
<tr>
<td>Serialists</td>
<td>Highlighted Keywords(Q1)</td>
</tr>
<tr>
<td></td>
<td>Lack of detailed instruction (Q2)</td>
</tr>
<tr>
<td></td>
<td>Hard to find the location of the Expert Search (Q3)</td>
</tr>
<tr>
<td>H6</td>
<td>Too many subject categories (Q17)</td>
</tr>
<tr>
<td></td>
<td>Excessive information in the</td>
</tr>
</tbody>
</table>
Table 4: Findings different from the Whole sample.

<table>
<thead>
<tr>
<th>H1</th>
<th>Highlighted Keywords (Q1)</th>
<th>Lack of detailed instruction (Q2)</th>
<th>Hard to find the location of the Expert Search (Q3)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Holists</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>H7</td>
<td>Only English version (Q20)</td>
<td>-.029</td>
<td>-.239</td>
</tr>
<tr>
<td></td>
<td>Provision of three different search mechanisms (Q21)</td>
<td>.507</td>
<td>-.354</td>
</tr>
<tr>
<td></td>
<td>Provision of three different types of font size (Q22)</td>
<td><strong>-.598</strong></td>
<td>.250</td>
</tr>
<tr>
<td><strong>Serialists</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>H6</td>
<td>Too many subject categories (Q17)</td>
<td>.542</td>
<td>.516</td>
</tr>
<tr>
<td></td>
<td>Excessive information in the Home page (Q18)</td>
<td>.083</td>
<td>-.194</td>
</tr>
<tr>
<td></td>
<td>Clear text icons (Q19)</td>
<td>-.559</td>
<td>-.289</td>
</tr>
</tbody>
</table>

Keys: * p < .05, ** p < .01

4. Conclusion

This study aims to examine how different cognitive style users respond differently to the interface design of the electronic journal. The major results of our research showed most of the students thought H8 was the most important heuristic. However, there are some differences between Holists and Serialists. More specifically, Holists who perceive excessive advertising may strongly need previous/next buttons while Serialists who feel this electronic journal provides too many advertising may feel too much information presented in the home page. Such differences between Holists and Serialists reveal that cognitive styles do play an important role. Accordingly, cognitive styles should be considered for the development of individualized mobile electronic journals. However, this study has several limitations. Firstly, the sample is small so further works need to use a larger sample to verify the findings presented in this study. Additionally, there is also a need to conduct further research to examine how other human factors, such as gender differences or prior knowledge, influence learners’ responses to the usability inspection of the electronic journals in the mobile context.

Acknowledgements

We would like to thank all the people who prepared and revised previous versions of this document.

References


Enhancing Metacognition through Weblog in Physics Classroom Thai Context

Jirutthitikan PIMVICHAI, Chokchai YUENYONG* and Sakanan ANANTASOOK

Science Education Program, Faculty of Education, Khon Kaen University, Thailand
* ychok@kku.ac.th

Abstract: This study examines the effects of using weblog technologies to create argumentative activities and enhance students express the metacognition on the situation's discussion. The Toulmin's diagram of Arguments Pattern (TAP) was adopted to interpret the students' metacognition in Thai rural physics classroom context. The participants were 33 Grade 10 students in the first semester of 2014 academic year, from a rural school in the Khon Kaen province of Northeast Thailand. The finding shows that rural students are very satisfactory to learn through the blog and the blog support them to post the comment for discussion. Their writing express metacognitive awareness, knowledge and control depend on each TAP’s coding schema. Consequently, this study provides effective weblog argumentative strategies for enhancing student’s metacognition in physics classroom Thai context.

Keywords: metacognition, argumentation, physics weblog

1. Introduction

A weblog (blog) is a web-based technology that is one of the popular web 2.0 teaching tools in this century (Namwar & Rastgoo, 2008). Blog can help to break down the classroom walls and increase students’ motivation to learn about science (Barlow, 2008). It is provide both teachers and learners with the ability to extend discussions away from the traditional face-to-face classroom. Blogs provide teachers as a user-friendly online format that can be used to emphasize strategies, introduce new topics, ask questions, review concepts, evaluate for tests, have argumentation and provide enhancement opportunities (Barlow, 2008). Students learn how to share their thoughts and communicate opinions which reflecting on real-world issue through comment on the blog (Duplichan, 2009). In addition, blogs can also give the “silent student” a voice, they are not very comfortable show their ideas in classroom. The blog can offer a “safe place” for their voices to be heard in a lower pressure environment (Luehmann and Frink, 2009). Students felt blog was “fun” and “helpful” and made them interested in using these technologies (Barlow, 2008; Columbo, 2007; & Erickson, 2009). It allows students the chance to take charge of their own learning (Luehmann and Frink, 2009).

Since 2010, The Technology for Teaching and Learning Center, under the Thai Ministry of Education every years training Thai teachers to create blog content through “wordpress” free blog service (URL: http://www.wordpress.com). Anantasook (2014) research study show that the students achievement after learning by using a weblog that applying social media entitle “basic C program language” was higher than before learning at a significance level of .05 and the satisfied of students who learning academic achievement with lesson on the blog in the high level as 4.29 of 5 score and standard based as 0.86. Another study suggests that teaching with lessons on the blog could decrease the differences between individuals. Students can study base on their own or preparing to study before attends in class. They can learn a lesson at anytime, anywhere when they comfortable and they can share ideas with friends via the comments box lessons. These can decrease the problem of students who not dare ask the teacher, answer the question or discussion in the classroom (Sudprakone, 2012).

According to the section 24 of the National Education Act of B.E. 2542 (1999) (Office of the National Education Commission; ONEC, 1999), provide the important idea of learning process, educational institutions shall provide training in thinking process, management, how to face various situations and application of knowledge for obviating and solving problems. The teacher must be aware of the importance of effective strategies for teaching and learning science (physics).
Physics is a scientific subject that is important to study in order to understand phenomena that occur and an introduction before study in higher education (Rosnow & Rosenthal, 1989; Aikenhead & Ryan, 1992). Most students do not like it, they think physics is a difficult subject, difficult to grasp and understand especially on the calculation content. Therefore, it must be improvements in the instructional model for teaching and learning factors that affecting the increase academic achievement. Yuenyong, Jones and Sung-ong (2011) suggested that metacognition is essential tool that supports students’ learning physics using higher-order thinking processes and situated cognition ideas are the arguments that show the influence of culture on students’ cognitive processes. When students learn how to create a scientific argument that there is a reason able to integration of the thinking skills with specific knowledge related to social issues. Students are able to use better reasoning in the support of the issue manually in order to promote the issues in argument are conflicting (Lin and Mintzes, 2010). They can develop confidence in making decisions in their lives and to participate as a responsible citizen in the social responsibility and democratic (Driver, et al., 2000).

However, from the previous researcher study revealed that Thai students from both urban and rural school are often shy, afraid to argue with friends and are afraid of express their own opinions in classroom. Particularly the students in rural school, they show a few comment and they do not attempt to describe the situation, discussion or arguments by their own. It seems like they lack metacognitive awareness, knowledge and control. From above advantage of weblog to support learning, the researcher interesting to create this technology as a metacognition strategy for enhance argumentative teaching and learning in rural physics classroom under Thai context.

2. Aims of the research

This research study examines the effect of weblog upon enhancing argumentation and metacognition in Thai rural physics classroom context.

3. Methodology

This study concern with qualitative approach, the findings stem from interpretive the students responded; how the effect of weblog upon enhancing argumentation and metacognition in Thai rural physics classroom context. The Toulmin’s diagram of Arguments Pattern (TAP) (Toulmin, 1958) were used to interpret the students’ metacognition that from six argumentation situations which shown in the comment wall on the blog.

Participants

The sample for the case study consisted of 33 Grade 10 students in the first semester of 2014 academic year that from Dongmon Wittayakom school, a rural small size school in the Khon Kaen province of Northeast Thailand.

Weblog as the instrument

The weblog [http://jirutthitikan.wordpress.com/] which concern in this study was created by the researcher. The objective's blog setting was provided for teaching and learning of argumentation in physics classroom at the force, work and energy Unit. This blog consist of five main menus; (1) Introduction of the Unit, (2) Lesson on the Unit, (3) Argumentation situation, (4) Participants, and (5) Mind Map.

The first introduction the unit menu includes the introduction on the physics area in Thai science curriculum and the objective of learning in the force, work and energy unit. The second lesson on the unit menu that include pre-post test and 6 contents of the force, work and energy unit which student must learn in class that are: (1) Force and Work (2) Energy, Kinetic Energy, Potential Energy (3) Law of Energy Conservation, (4) Apply Law of energy Conservation, (5) Power, and (6) Mechanical. The third argumentation situation menu, that include six argumentation situations which relevant with six
contents in the previous menu. The fourth participant menu that provide the participant information in this study and the final mind map menu will show students’ task after they finish learning the unit.

4. Data collection and data analysis

Five students’ groups completed argumentation in the six argumentative questions with their friends. They also express evidence of their conceptual understanding “thinking of thinking” (metacognition) by post or comment discussion. The metacognition on student discussion could found through Toulmin’s diagram of Arguments Pattern (TAP) (Toulmin, 1958). It consists of 6 categories that include: Data, Claim, Warrants, Qualifiers, Rebuttals and Backing that shown in Table 1. The interpretive of students’ argumentation typically looks at three dimensions of metacognition (metacognitive awareness, knowledge and control).

Table 1: Coding scheme from Toulmin's diagram of Arguments Pattern (TAP)

<table>
<thead>
<tr>
<th>Discourse Move</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Data</strong></td>
<td>Students can use facts or evidence to prove their argument.</td>
</tr>
<tr>
<td><strong>Claim</strong></td>
<td>The statement being argued. Students’ principle comment or an assertion made by students’ brainstorm in groups. Arguments which are a simple claim versus a counterclaim or a claim versus a claim. Argument which may have several claims and counterclaims.</td>
</tr>
<tr>
<td><strong>Warrants</strong></td>
<td>A student has arguments consisting of a claim versus a claim with data, warrants or backings but do not contain any rebuttals. The general, hypothetical (and often implicit) logical, statements that serve as bridges between the claim and the data.</td>
</tr>
<tr>
<td><strong>Qualifiers</strong></td>
<td>Statements that limit the strength of the argument or statements that propose the conditions under which the argument is true.</td>
</tr>
<tr>
<td><strong>Rebuttals</strong></td>
<td>Counter arguments or statements indicating circumstances when the general argument does not hold true. Sometime students are subject to argumentation displays and extended argument with more than one rebuttal or argumentation has arguments with a series of claims or counterclaims with data, warrants or backings with the occasional weak rebuttal. Although argumentation shows arguments with a claim, with a clearly identify able rebuttal.</td>
</tr>
<tr>
<td><strong>Backing</strong></td>
<td>Statements that serve to support the warrants (i.e., arguments that don't necessarily prove the main point being argued, but which do prove the warrants are true).</td>
</tr>
</tbody>
</table>

According to Table 1, TAP reflects the students’ thinking that are considered in the metacognition includes: 1) Metacognitive awareness: students know, use, and can interpret scientific explanations of the methodology of force, work and energy; 2) Metacognitive knowledge: students understand the nature and development of scientific knowledge of force, work and energy; and 3) Metacognitive control: students can generate and evaluate scientific explanations and arguments of force, work and energy.

5. Finding and Discussion

The dialogue of students’ discussion in six argumentative situations was shown in the blog. The information was then searched to identify episodes of metacognition from dialogical argument. Dialog of the rural students who attend in the discussion or debate comments on various issues of argument on the blog was frequent and there also diverse missed opportunities for giving the quality of the feedback to students. The evidence on all aspects of teaching through blog investigated that the arguments are relied on the use of metacognition for discussion that shown in each component of Toulmin’s diagram of Arguments Pattern (TAP) as the following:
5.1 Data (D) Component

The students’ comment during weblog learning activities provided the data corpus on the students’ argument engagement. Even though, some answer is correct, the claim or explanation of situation is not enough that could be found in the dialogue of the Law of Energy Conservation content in table 2.

Table 2: The dialogue from the comment blog wall about the Law of Energy Conservation content.

<table>
<thead>
<tr>
<th>Teacher</th>
<th>We are studying the law of energy conservation. This is the question: Throw up an object at the position X vertically until it reaches the highest position Y. Then, the objects drop down to position Z (Figure). Compare the kinetic energy of the object at the position X, Y, and Z?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group 1: Student 3</td>
<td>I think Y has Kinetic Energy (Ek)=0 because the velocity (V_y)=0</td>
</tr>
<tr>
<td>Group 2: Student 1</td>
<td>I think when the object dropped, its velocity increases until it has constant velocity (V) at the position Z.</td>
</tr>
<tr>
<td>Group 3: Student 2</td>
<td>I think the velocity at X is higher than at Z (V_x&gt;V_z) and velocity Y is Zero (V_y=0)?</td>
</tr>
<tr>
<td>Group 4: Student 3</td>
<td>I think the highest kinetic energy stay at Z and zero energy at Y.</td>
</tr>
<tr>
<td>Group 5: Student 1</td>
<td>I think the lowest kinetic energy stay at Y and Z more than X position</td>
</tr>
</tbody>
</table>

The students’ scientific conception, scientific attitude, logic and reasoning come from their daily context and previous experience in the study of self-knowledge. Students usually use the data from textbook and the content on the blog in an argument with friends in the classroom. The teachers motivate students to participate in learning activities using the score as encourages them to raise the learners’ concentration in classroom. Therefore, it makes students eager to find out the comment to get the most points.

Data (D) Component and the metacognition

The question and answers are carefully chosen to reflect the relationship between the teacher and students in the physics blogging classroom using “Data” component found that; students have used the fact to prove the comment of argument. The student’s uses data which articulate and expand the idea in each comment. They post their ideas from other sources than the main blog to refer to their claim. In this situation, the role of students is very highly use metacognitive awareness, knowledge, control because they had many source of data in order to answer the question.

5.2 Claim (C) Component

In the claim component, students have few claims on all questions because some groups of students tend to agree with the claims of friends. Sometimes they would not dare to claim on their own; students do not want to comment the questions themselves because they fear of giving wrong comments. Some comment like as agree with their friend that rarely show the scientific explanation. Therefore, students will agree with the conclusion that the majority believes in and assert it as the true claim. Sometimes students may have several claims and counter claims. Students can claim by explaining the reasons with data and claims attributed to comment questions. Teachers compared the claims or comments of each group at the end of the argument to conclude which of them is correct. This could be seen in the following dialogue: Apply the Law of energy conservation (table 3).
Table 3: The dialogue from the comment blog wall about Apply the Law of energy conservation content.

Teacher What kind of energy is the bungee jumper involve in? Describe it at each of the following stages:

A. Before Jumping
B. While the bungee cord isn’t stretched
C. While the bungee cord is stretched the most

Group 3 (Discussion in group)
- I think (A) has only gravitational potential energy of jumper.
- I think same too! and then what do you think about B?
- Hmm !!! Figure (B) if viewed in the figure B “I think in the first period, gravitational potential energy of jumper is decreased, kinetic energy is increased.” Anyone agrees with my idea or not?
- I agree it should be possible and figure (C)? Who can explain and how do you know?
- Figure (C) “I think this period has the elastic potential energy of bungee cord and gravitational potential energy of jumper.
- And then how will you explain about gravitational potential energy.
- (C) this period has the elastic potential energy of bungee cord and gravitational potential energy of jumper. Everyone agrees! Ok?

Group 3 conclusion (claim)
- We’re thinking (A) has only gravitational potential energy of jumper, (B) in the first period, gravitational potential energy of jumper is decreased, kinetic energy is increased and (C), this period has the elastic potential energy of bungee cord and gravitational potential energy of jumper.

Claim (C) Component and the metacognition

The questions and answers chosen to reflect on the relationship between teachers and students in the physics blogging classroom using “Claim” component found that; the students’ arguments in blogging classroom are interesting. Students comment on each issue with many claims in the force, work and energy content while some groups, students tend to agree with the claims of their friends instead of making claim themselves. As a result, the class mocked or blamed the situation so that the student is the audience more than the claimer. In the dialogs of claim component, students’ metacognitive awareness, knowledge and control are very slightly use by owns self. They are flowing groups and teachers more than trust owns self.

5.3 Warrants (W) Component

In the warrants component, when students are into claim and discussion, Students often have evidence for the claim and are used for reliable data to support their comment questions that are about the topic. The students when it is claimed by warrants it consists of reasoning, assumptions and source of comment. Some students are drawing their own on the book and show in class (it is difficult to post the picture in the blog at that time) to warrants their claims or may be more of the comment to other groups. This can be seen from the following dialogue about force and work content (table 4).
Table 4: The dialogue from the comment blog wall about force and work content.

<table>
<thead>
<tr>
<th>Teacher</th>
<th>Pushing a toy car on a flat surface (case A) and push it up the slope (case B) with equal speed, distance, and acting force. Which one use more work? And why?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group 4</td>
<td>We’re think case B cost more work than case A; FB &gt; FA”</td>
</tr>
</tbody>
</table>

Group 4 Student 2  A student affirms the claims by drawing (that shown in the figure).

**Warrant (W) Component and the metacognition**

The questions and answers chosen to reflect on the relationship between teachers and students in the physics blogging classroom using “Warrant” component found that; when students prepare to make claim and discussion, they often have evidence for the claim and use reliable data to support their comments. The students claimed that warrants consist of reasoning, assumptions and source of comment. Some students draw pictures by themselves to warrant their claims. The comment on this physics arguments, most students would agree with or justify, combine ideas organize the discussion. Students’ metacognitive knowledge in the warrant component was shown more than metacognitive awareness and control. Student usually used knowledge from their old experience and the teacher teach through a blog to support their reason.

**5.4 Qualifiers (Q) Component**

The results from six argumentative activities indicated that students use qualifiers that propose the conditions under which the argument is true in physics blogging classrooms rather than propose the limitation of their friend’s arguments. In this dialog of qualifiers component, students’ metacognitive awareness, knowledge and control couldn’t shown metacognition since they are not aware of qualifiers for decide to answer the argumentative question. The students’ debate based on their experience and teachers though then they never think of the condition or limitation while discussion in blogging classroom. However, when students need to use picture explanation, they wrote on the book and they could not post on the blog. It could be mention that shown the limitation of arguments via the blog.

**5.5 Rebuttals (R) Component**

In the rebuttals component, students agree rather than rebut with the hero who can explain or answer questions. Most students who are not proficient enough or have not experience usually consider their friends as heroes without rebuttals. They are maybe afraid to comment because they have that notion that the comment which receives rebuttals is wrong. The student replied that if the comment is wrong, everyone can saw that is a shameful thing for them. This can be seen from the following dialogue about power content (table 5).

Table 5: The dialogue from the comment blog wall about power content.

<table>
<thead>
<tr>
<th>Teacher</th>
<th>Could you try to solve this problem? The different way to the waterfall at the top of mountain; first is tortuous way and less slope, another is a straight line but more slope. Which way are easy and spent little time?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group 2 Student 1</td>
<td>I think the first way is easier because it has less inclined slope and uses lesser power.</td>
</tr>
<tr>
<td>Group 2 Student 2</td>
<td>I agree with you. Anyone have any answer other than that? (No comment from other)</td>
</tr>
<tr>
<td>Group 2</td>
<td>Claim that “I think the first way is easier because it has less inclined slope and uses lesser power.”</td>
</tr>
</tbody>
</table>
Table 5: The dialogue from the comment blog wall about power content (continue).

<table>
<thead>
<tr>
<th>Group 3 Student 1</th>
<th>I think the first way is easier and save energy because the driver can slow drive. While the second way need high accelerate for release the high friction force.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group 3 student 2</td>
<td>I agree with you. Anyone have any answer other than that? (No comment from other)</td>
</tr>
<tr>
<td>Group 3</td>
<td>Claim that “the first way is easier and save energy because the driver can slow drive. While the second way need high accelerate for release the high friction force.”</td>
</tr>
</tbody>
</table>

Rebuttals (R) Component and the metacognition

Students rarely make rebuttals to the argument, because most students who are not proficient often wait for ideas from their “hero” friends. They are sometime afraid to comment because they consider an argument that has rebuttals is wrong. The student replied that if the comment is wrong, everyone can seen that is a shameful thing for them and may be they thought that their friends make appropriate explanation. Therefore, students show a little of metacognitive control because they do not like arguments to express their idea, they usually wait scientific explanations from their “hero” friends. However, students who can debate that mean they expressed more metacognitive awareness and metacognitive knowledge which both kind of metacognition appropriate for rebuttals.

5.6 Backing (B) Component

In the backing component, students were subjected for argumentation and had arguments with a series of claims with data, warrants and backings obtained from the references of the theory of energy (kinetic energy and potential energy). Even though reference is a reliable reason to support the comment of them, the most students tend to support ideas from a group of other friends rather than their own comments. Sometime they tend to wait for the hero group started comment and followed it although the answer was incorrect. This can be seen from the following dialogue energy content (table 6).

Table 6: The dialogue from the comment blog about energy (kinetic energy and potential energy) content

<table>
<thead>
<tr>
<th>Teacher</th>
<th>A and B carry the basket with the same size and weight. A climbing up the vertical ladder, B climbing up the slope ladder. At the same height, who make higher energy on the basket?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group 2</td>
<td>The energy to carry the basket A and B up to the top have to be equal because they are at the same height level. (After students discussion in their group and start to share with other group)</td>
</tr>
<tr>
<td>Teacher</td>
<td>“Group 2 comment that the necessary energy to carry the basket A and B up to the top have to be equal because they are at the same height level” Do you have anything different from group 2 idea?</td>
</tr>
<tr>
<td>Group 3</td>
<td>Group 3 agree with Group 2 because energy of the basket A and B are potential Energy (Ep). Due to Ep= mgh the question fix the same m, h, and g, then EpA=EpB</td>
</tr>
<tr>
<td>Group 1</td>
<td>Group 1 agrees with group 3 and group 2 because energy of the basket A and B were caused by additive height from potential energy (Ep).</td>
</tr>
<tr>
<td>Group 4</td>
<td>Group 4 agrees and support the claim from group 1 because of the equal height and equal weight due to EpA=mgh; EpA=EpB</td>
</tr>
<tr>
<td>Teacher</td>
<td>Concluded that the questions of A and B on the wall, these two cases are the same height. Energy of the two baskets will both increase equally with the equation mgh.</td>
</tr>
</tbody>
</table>
Backing (B) Component and the metacognition

Students are subject for argumentation and had arguments with a series of claims with data, warrants and backings obtained from the references of the theory. Even though reference that is a reliable reason to support their comments. In this “backing” component, it is found that most students tend to support idea from their friends’ group rather than their own. Some students tend to wait for another group makes the comments. Sometime they tend to wait for another group, which is a hero group that usually comments, to give ideas and agree with them. From these situation, it seem like students show a little metacognitive awareness, knowledge and control. They do not attempt to give describe the situation or comment by their own.

6. Conclusion and Suggestion

This research study revealed that learned by weblog which provide the scientific content and the argumentative activities could be motivate students to express their own ideas, opinions on situation’s discussion. It expressed the students’ metacognitive thinking on the situations that demonstrate on the post wall. The rural students claim on various issues commonly learning in the class. The students claim (C) on the Data (D) and warrants (W) component. Other elements comprise of the rebuttals (R) and backing (B) in which students made was based on their experience and content knowledge from the blog and their teacher. For the qualifiers (Q), students usually do not claim in each contents. In metacognition dimension, students express very highly metacognitive awareness, knowledge and control in the use many source of data (D) for debate. While they had very slightly metacognitive awareness, knowledge and control for claim (C) their ideas. Students’ metacognitive knowledge in the warrant (W) component show more than other kind of metacognition. In addition, students who can debate that mean they had metacognitive awareness and metacognitive knowledge for rebuttals (R). However, students do not show metacognition in the qualifiers because they never think of the condition or limitation while discussion in blogging classroom.

Interestingly, students prefer to post the comment for discussion, this maybe they never learn with the blog, learning through blog is the new approach for them. Learning on the blog could encourage students to higher thinking from argumentation and it could reduce the problem from students who not dare ask teacher, they could share the understanding on learning with friends via the comments box lessons. However, sometime they are aware that their comment maybe wrong, that make them more serious and lack the confident before share some ideas on the blog. In addition, the study also reveal that students who learned through the lesson on web blog very satisfactory. They were exciting and funning with learning because lessons’ blog are interesting, easy to use and quickly access lessons.

These findings suggest that the teacher can create the scientific content blog within the argumentative activities in Physics for enhancing student’s metacognition in Thai context.

7. Acknowledgements

This research study was supported by the Higher Education Research Promotion and National Research University Project of Thailand, Office of the Higher Education Commission, through the Cluster of Research to Enhance the Quality of Basic Education.

8. References


Knowledge Propagation in Practical Use of Kit-Build Concept Map System in Classroom Group Work for Knowledge Sharing

Toshihiro NOMURA, Yusuke HAYASHI, Takuma SUZUKI & Tsukasa HIRASHIMA

Abstract: This study proposes a design of class in which learners collaboratively organize and share what they learned in lessons. Concept mapping is a well-known technique that can help students to create visual representations of the structure of their understanding. This study uses a special kind of concept mapping for knowledge sharing, Kit-Build method. The characteristic of this method is to provide parts to build a concept map for learners. A teacher makes the parts and thus he/she can manage what students should learn and assesses answers automatically. This paper reports the result of classes conducted with Kit-Build concept map and analyzed the knowledge propagation in the collaborative learning. Through collaborative Kit-Build mapping among students the concordance rate of learner maps went up and the score of the maps improved. This indicates that students have shared and corrected their knowledge through the activity. In this study the teacher could realize the understanding of students and taught them what did not understand well based on the feedback of KB map system.

Keywords: kit-build, concept map, collaborative learning

1. Introduction

Development of creativity is one of the important goals in education (Griffin, 2012). On the other hand, Jonassen (Jonassen, 1992) and Scardamalia et al. (Scardamalia, 2012) point out a stage prior to develop creativity. Jonassen defines three stages of knowledge acquisition; introductory, advanced, and expert. At the introductory stage, learners acquire the knowledge that is in well-structured domains and it has only a correct answer for a question. At the advanced stage, learners acquire more advanced knowledge, that is in ill-structured domain and it could have various answers for one question. Expertise is the final stage where experts have more internally coherent and more richly interconnected knowledge structures. Scardamalia specifies two levels of knowledge construction; entry and high level. These levels are the equivalent of Jonassen’s introductory and advanced stage, respectively.

This study focuses on knowledge construction for the basis of creative tasks. The aim of this study is to design learning activities in classroom intended to encourage knowledge construction prior to development of creativity. In this study, according to Jonassen and Scardamalia et al., the requirements for the activities are that students can build their own knowledge inductively and check correctness of it in interaction with other students. They need to correct their knowledge as necessary and finally they share correct knowledge. This shared and correct knowledge is expected to lead their next activities for development of creativity. From the viewpoint of the definition of collaborative learning by Dillenbourg (Dillenbourg, 1999) the required elements are symmetry of action and status of students, interaction among them and inductive thinking.

To satisfy the requirement this study adopts collaborative learning with Kit-Build concept map (Yamasaki, 2010). Kit-Build concept map is a kind of closed concept map construction system in which learners make concept maps from parts provided from a teacher. Although this is similar to Expert skeleton concept map, the difference from it is that Kit-Build concept map requires learners to completely rebuild a concept map from fragmented pieces of concepts and links. Learners relive the consideration of relation among concepts in the target domain. This method is also effective in diagnosis of learners’ understanding. The concept map represents the structure of what the teacher want
students to understand. A learner's understanding can be evaluated as the concordance rate between a map made by the learner and one made by the teacher. General concept mapping uses open system in which there is no restrictions to build concept maps and constructors use any nodes and links. In this case, it is difficult to diagnose concept maps (Herl, 1999). There are studies show the effectiveness of Kit-Build concept map in evaluation of and feedback to learners based on it (Sugihara, 2012) (Yoshida, 2013).

Maldonado et al. proposes a system to support this individual and collaborative creative knowledge building by using concept map and ICT literacy (Maldonado, 2012). He tracked and analyzed the flow of knowledge that is created as a result of individual pre and post-concept maps at the personal computer and group concept maps construction at the multi-touch tabletop. As mentioned above, ICT literacy can offers the possibilities to extend the support to students and moreover can be an analytical tool by which teachers support students. Also, the purpose of this study is to support knowledge building by using ICT literacy. However, our study intends to different level from his study.

This paper reports practical use of Kit-Build concept map system in classroom group work for knowledge sharing and knowledge propagation among them in the learning. The remainder of this paper is organized as follows. The next section explains Kit-Build map system that is the core technology of this study. Section 3 presents the design of classes the authors conducted. These classes are conducted in a junior high school as a part of usual classes by one of the co-author. Section 4 shows the result of the class and analysis of the data. Last section concludes this paper.

2. Kit-Build Concept Map

Concept maps (Novak, 2006) are graphical tools for organizing and representing knowledge or understanding. A concept map includes concepts and relationships between concepts indicated by a connecting line linking two concepts. Two concepts linked with a relation represent a proposition. Concept map is effective for facilitating learning and for enabling learners to create visual representations of their comprehensive structure. This has enormous significance in enabling to evaluate and share learners’ knowledge.

Kit-Build concept map (KB map) is a framework to build and diagnose concept maps (Yamasaki, 2010). In this method, learners build concept maps (learner map) by assembling provided parts. The parts are generated by decomposing an ideal concept map (goal map) that is prepared by a teacher as the goal of his/her teaching. The characteristic of KB maps is that it is possible to compare and overlap them because both of the maps, learner maps and goal maps are composed of the same components. Comparing them, system can calculate the concordance rate of the map as the score of learner maps automatically. KB map provides the teacher with information about learners’ comprehensions and differences of their thoughts on the map that overlaps all the learner maps (group map). Also, because of providing the same parts among learner and goal maps, it is easy for students to compare their maps in discussing and negotiating understanding with maps and KB map clarify the differences between a learner’s and others’ understanding in creating the map collaboratively.

Figure 1. KB map editor.
There is a system to realize interaction based on Kit-Build method. This system is called KB map System. This system is a web application with two client systems: KB map Editor and KB map Analyzer. KB map Editor provides an environment to make a learner map. This system can be used with tablet pc (Sugihara, 2012). Figure 1 show the picture of the KB map editor. This permits students to carry their maps and to show their maps to the other students for discussing. KB map Analyzer has function to overlap learner maps and teachers understand his/her students’ comprehension with this system.

3. Outline of Practice in a Junior High School Japanese Social Studies Class

3.1 the Purpose and the Design Principle of the Class

This study designed and conducted the lessons in which students use KB map. This is collaboration with the junior high school teacher that is one of the authors of this paper. He has a desire to make students enhance their understandings with both of collaboration and teaching.

In the lessons what students do is the following two things: to organize their own knowledge on KB map inductively from materials and what they have learned in the previous lessons and to compare and correct their knowledge represented on KB map through discussion. After that, the teacher explains the correct answers compared with students' KB maps. Through this process, this lesson expects that students having incorrect knowledge learn correct one from the others and that the teacher identifies students' misunderstandings still remained after discussion and teach them carefully.

3.2 Design of the Class

In this study, the procedure of the class was the following: Firstly the teacher does a review of the topics in the previous classes. The teacher shows some pictures related topics and explain it. Secondary, students make KB map (pre-map) individually. In this phase they organize their knowledge inductively from materials and what they have learned in the previous lessons. Thirdly, they go into a small group and work together. They discuss the difference among their maps and make a KB map of the group (collaborative-map) collaboratively. Fourthly, they make modify their own KB map (post-map) individually again if necessary. Lastly, the teacher gives feedback about the topic with the collaborative-maps students have made. In this procedure, pre-maps underlie collaborative-maps, and post-maps reflect collaborative-maps. The procedure of the class is shown in Figure 2.

![Figure 2. The Flow of This Class and the Data from the Maps.](image-url)
3.3 Hypothesis

In this research, three hypotheses were formed. These are the followings:

Hypothesis 1: After group work, students have much more common understanding about the topic than before. (in a group the similarity of post-maps is higher than the one of pre-maps)

Hypothesis 2: After group work, students have much more correct understanding than before. (post-map scores are higher than pre-map scores)

Hypothesis 3: The teacher can realize the understanding of students based on group map that is made by KB map system.

3.4 Participants and Procedure of the Practice

Three lessons were conducted for three first grade classes in a junior high school. The participants are 76 Japanese students who are 12 or 13 old in total from three classes. The number of students in the class A is 26, in the class B is 25 and in the class C is also 25. In the classes, the teacher regularly uses concept maps for teaching social studies and thus, the students were used to make concept maps in learning it. In addition to that, they also have ever used KB map system in studying social studies lesson. Therefore, it is not so difficult for them to make concept maps and to use KB map system.

The subject domain of the class conducted in this study was the unit of South America in geography. The theme was the dilemma of economic development and deforestation. The students had mainly studied economic development in South America and had had knowledge about deforestation within the bounds of common sense.

In this practice, firstly the teacher makes a concept map about industrials and traffics in South America on the blackboard with the students. This map making is a task to recall what they studied in the previous classes. They made the parts of the map in several lessons. Secondly the teacher provided tablet computers with students. The kit of this class is shown in Figure 3. With a tablet, each student made a KB map including the map on the blackboard. The part of the map that is on the blackboard is already constructed on the KB map and the students are required to link the rest of it with the separated concepts and links. The separated concepts are “economic development” and “deforestation.” The students are required to consider which concepts connect to economic development and/or deforestation. The kit includes enough unfixed links to connect all the concepts about industrials and traffics with both of them. In the goal map, all the separated concepts connected to both of economic development and deforestation. However, the teacher did not clearly tell the students that they did not have to use all the links. Therefore, the students must have considered whether each concept was connected with economic development and/or deforestation or not. In the previous class, the students

![Figure 3. Kit of This Class.](image-url)
had learned economic developments and industrial, however, they had not learned deforestations and traffics. The teacher gave the students some documents about deforestations, and students could refer it when they made KB map. Thus, this assignment consisted of both recall and generation. Making the connection to economic development is a recall task because the teacher explained in the previous lessons. On the other hand making the connection to deforestation is a generation task because they need to consider with their pre-existing knowledge and the provided pictures. As the first step of the map making, the students make a map individually as a personal opinion. This map is called “pre-map,” here. After that, they went into a small group of four or five students and discussed the difference among their maps and made a map collaboratively as an agreement. This map is called “collaborative-map,” here. Next, they could modify their map individually if they need after group work. This map is called “post-map.” Finally, the teacher explained the correct answer in comparison with the collaborative-maps. Here, the teacher used the group map that was made by overlapping every collaborative-map.

4. The Result and Consideration

4.1 Testing the Hypotheses

Hypothesis 1: After group work, students have much more common understanding about the topic than before (in a group the similarity of post-maps is higher than one of pre-maps)

Figure 4 and table 1 show the comparison among the average concordance rates between the pre- or post-maps in each group. It is calculated by the following equation:

\[
\frac{\text{the number of the pairs of students who connected the same link to the others}}{\text{the number of the propositions in goal map}} \times \frac{\text{the number of group members}}{\text{the number of the propositions in the goal map}}
\]

The average concordance rate among post-maps is higher than the one among pre-maps. There is a significant difference between them in every class (in the class A: two-sample t-test, \(t(5) = 4.8906, p<0.01\), in the class B: two-sample t-test, \(t(5) = 4.0865, p<0.01\) and in the class C: two-sample t-test, \(t(5) = 6.3571, p<0.01\)). This indicates that students shared their understanding and built a consensus in each group in some way.

Furthermore, Figure 5 and table 1 show the comparison among the average concordance rates between pre- or post-maps and the collaborative-map in each group. This is calculated by the following equation:

\[
\frac{\text{the number of the propositions that are common between the individual and collaborative map}}{\text{the number of the propositions in the goal map}}
\]

The concordance rate between post- and collaborative-maps is higher than the one between pre- and collaborative-maps. There is a significant difference between them in every class (in the class A: Wilcoxon signed-rank test, \(n=26, V=228, p<0.01\), in the class B: Wilcoxon signed-rank test, \(n=24, V=153, p<0.01\), in the class C: Wilcoxon signed-rank test, \(n=25, V=253, p<0.01\)). This indicates many students changed their individual-maps following their collaborative-maps after discussion.

Hypothesis 2: After group work, students have much more correct understanding than before (post-map scores are higher than pre-map score).

Figure 5 and table 2 show the comparison between the average scores of the pre-maps and the post-maps. A score of the learner map indicates degree of similarity between the learner map and the goal map. It takes the value of 0 to 1. If the score is 1, it means the learner map is completely same as the goal map. The score is calculated by the following equation:

\[
\frac{\text{the number of the correct propositions in a learner map}}{\text{the number of the propositions in the goal map}}
\]
The average scores of the post-maps are higher than ones of the pre-maps in every class. There is a significant difference (in the class A: Wilcoxon signed-rank test, \(n=26, V=210, p<0.01\), in the class B: Wilcoxon signed-rank test, \(n=24, V=148.5, p<0.01\), in the class C: Wilcoxon signed-rank test, \(n=25, V=253, p<0.01\)) between them. This indicates that the understanding of students was improved through making the KB map in a group.

Moreover, this study compares the score of the pre-map to the score of the collaborative-map using the Wilcoxon rank sum test. A significant difference was found between the score of the pre-map and the score of the post map in every class (in the class A: \(n=26, U=31.5, p<0.05\), in the class B: \(n=24, U=34, p<0.01\), in the class C: \(n=25, U=34.5, p<0.01\)). The scores of the collaborative-maps are higher than the scores of the pre-maps. This finding suggests that the knowledge created collaboratively is more correct than the average knowledge created individually.

Consequently, because individual learner maps came closer to collaborative-maps, the degree of the similarity among group members’ maps increased, and because the score of the collaborative-map was higher than the score of the pre-map, the students’ knowledge improved.

Consequently, the summary of the results is the followings:
(1) the concordance rate of individual-maps in groups increased through group work, and
(2) the concordance rate between individual-maps and collaborative-map also increased through group work, and
(3) the score of individual-map is improved.

These can be considered as the followings:
(1’) the students share their understanding in their group, and
(2’) they made a collaborative-map as their shared understanding, and
(3’) their shared understanding improves their individual-understanding.

The authors carried out a questionnaire survey to clarify how students had made the collaborative-maps and what is the reason if students had changed their maps after group work. This questionnaire includes two questions about decision making in group work and improvement of individual-map after it: “How did you make your collaborative-map?” and “How did you change your individual-map?” Students could choose from three options: “(1) by majority vote”, “(2) agreement on the others’ opinion” and “(3) by just following others’ opinion”. The 60-70 percentages of the students chose the second option on both questions. This suggests that students place importance on the agreement with the opinion by discussing when they change their opinion and make their collaborative knowledge. In addition to that, the relation between pre-maps and collaborative-maps shows data supporting the suggestion. Table 3 indicates the relation between pre-maps and collaborative-maps. In each group, the proposition in the pre-maps is categorized according to the number of the students that they have the same propositions in their pre-maps. Moreover, the propositions are categorized by the correctness of the propositions in their pre- and collaborative-maps. Note that students did not always choose the proposition by majority vote. Of course, they chose by majority vote, however, there are some cases that they chose from minority.

Hypothesis 3: The teacher can realize the understanding of students based on group map that is made by KB map system.

In this class, the teacher made a group map from the collaborative-maps, and compared the group map with the goal map to show how many groups did not make the correct links. Figure 7 is the group map composed of the lacking links in the collaborative-maps this study obtained in the class-A. According to the group map, the teacher could get the information about the propositions that the groups had a lack of understanding toward and were divided on the opinion of. The teacher supposed that the students had lack of understanding toward the propositions about the new concepts, railways and roadways. However, according to the group map, there was the lack of understanding toward the various propositions that the teacher did not suppose. The greatest numbers of the propositions that students had the lack of understanding about were, in class A “Forestry relate to economic developments.” and “Factory relate to deforestations.”, in class B “Farm relate to economic developments.” and “Farm relate to deforestations.”, in class C “Forestry relate to economic developments.” and “colliery relate to deforestations.”. In this class, the teacher could explain mainly about these propositions. After the lessons the teacher said that this is the first time to get information about understanding of students on time in classroom. Although he had tried to investigate individual thought of students during group work, it needed the help of other many teachers and it is difficult to organize
the result during the lesson. This time he satisfied the information and give instruction based on it during the lesson.

![Figure 4](image1.png)  
**Figure 4.** The Average Concordance Rate among Individual Maps in the Groups.

![Figure 5](image2.png)  
**Figure 5.** The Average Concordance Rate between the Individual and Collaborative Map.

![Figure 6](image3.png)  
**Figure 6.** The Average Map Scores in three classes.

<table>
<thead>
<tr>
<th>Table 1: Concordance Rates.</th>
<th>Pre-maps in the group</th>
<th>Post-maps in the group</th>
<th>Pre-Collaborative</th>
<th>Post-Collaborative</th>
</tr>
</thead>
<tbody>
<tr>
<td>Class-A</td>
<td>0.7104(SD=0.0399)</td>
<td>0.8661(SD=0.0462)</td>
<td>0.6346(SD=0.2170)</td>
<td>0.8389(SD=0.1563)</td>
</tr>
<tr>
<td>Class-B</td>
<td>0.7031(SD=0.0733)</td>
<td>0.8724(SD=0.0858)</td>
<td>0.6198(SD=0.1861)</td>
<td>0.8255(SD=0.1403)</td>
</tr>
<tr>
<td>Class-C</td>
<td>0.7708(SD=0.0696)</td>
<td>0.9563(SD=0.0296)</td>
<td>0.6650(SD=0.2064)</td>
<td>0.9200(SD=0.0875)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 2: Map Scores.</th>
<th>Pre-map</th>
<th>Collaborative-map</th>
<th>Post-map</th>
</tr>
</thead>
<tbody>
<tr>
<td>Class-A</td>
<td>0.5817(SD=0.2206)</td>
<td>0.8125(SD=0.1936)</td>
<td>0.7764(SD=0.1922)</td>
</tr>
<tr>
<td>Class-B</td>
<td>0.5885(SD=0.2395)</td>
<td>0.7708(SD=0.1407)</td>
<td>0.7839(SD=0.1758)</td>
</tr>
<tr>
<td>Class-C</td>
<td>0.6400(SD=0.2226)</td>
<td>0.8958(SD=0.1164)</td>
<td>0.8950(SD=0.0967)</td>
</tr>
</tbody>
</table>
Table 3: the Relation between the Pre-Maps and Collaborative-Maps.

<table>
<thead>
<tr>
<th>The type of result</th>
<th>Unanimous accord</th>
<th>Majority</th>
<th>Even</th>
</tr>
</thead>
<tbody>
<tr>
<td>Which is common in pre-maps</td>
<td>Correct agreement</td>
<td>Incorrect agreement</td>
<td>Major correct opinion</td>
</tr>
<tr>
<td>Correct or Incorrect in collaborative-maps</td>
<td>C</td>
<td>I</td>
<td>C</td>
</tr>
<tr>
<td>The number of Propositions</td>
<td>51</td>
<td>2</td>
<td>3</td>
</tr>
</tbody>
</table>

4.2 Analysis of the Knowledge Propagation in Collaborative Learning

As mentioned above, the students’ knowledge improved on the whole. However, this result is just average. Now let’s take a look at the several students. The students did not improve all of their misunderstanding. Although, some students corrected their misunderstanding of some items through group work, other students went wrong in understanding of some items after group work even though they had had correct understanding. About these phenomena, this study examined the knowledge propagation in more detail.

Table 4 shows the patterns of change of understanding and the distribution of them. There are eight patterns of change among pre-, collaborative- and post-maps by the correctness of the propositions in each map. The distribution is derived from about tallied the 1200 propositions in the pre-, collaborative- and post-maps.

The most common pattern is Pattern-A: students had had correct understanding of a proposition from the beginning and then had kept it at the end. This pattern accounts for over half of all propositions. Because the average score for the pre-map was about six out of ten and it improved in post map, this may be the reason why their map score was kept.

The second most common pattern is Pattern-E: students had had incorrect understanding of a proposition at the beginning and then they changed it to correct one following the agreement in their group. This pattern accounts for over two out of ten propositions and accounts for about half of the propositions that the students got the wrong answer about in their pre-maps. This may provide significant share of the reason why their score for pre-map improved.

The third most common pattern is Pattern-H: students had had incorrect understanding of a proposition at the beginning and then they had kept it by the end. The agreement in their group is also incorrect, therefore, they did not have chance to change their understanding. This pattern accounts for over one out of ten propositions. Most of this case happened when no one had had the correct answer in the group. This suggests that it was hard for students to improve their understanding without the member who had had the correct understanding. Especially, the KB map in this study requires students to make an either-or decision. Therefore, this does not produce the diversities of opinion. The students
often reached the agreement on the wrong answer from the beginning. It seems that the students did not discuss the propositions in this situation. A similar result was found in another study with the KB map.

These data supports the result of hypotheses testing in the last section. Although the above patterns are majority, there are some other patterns. In some patterns the students changed their answer from correct one to incorrect one in the post-map.

In Pattern-C and Pattern-F the students did not follow the decision in their collaborative-map and kept their own understanding that is different from the group’s decision. In Pattern-D the students changed their correct understanding to incorrect one following the incorrect group decision. The number of Pattern-C is twice the number of Pattern-D. This may suggest that students prefer to keep their correct answer than to get groups’ wrong answer.

Pattern-G and pattern-B mean that although a proposition in the pre-map was same to the propositions in the collaborative-map they change the proposition in the post map. They changed their thought after group work. However, there is no data to identify the reason in this study. In order to identify it, it is necessary to gather data about what they talked about in the group work other than data about KB map.

Regarding the propositions that changed from the pre-map to the post map, such as pattern-B, -D, -E and –G, nine out of ten propositions changed from incorrect to correct. Moreover, in this case, nine out of ten propositions changed to the propositions that were made collaboratively.

Table 4: the Patterns of Change of Understanding.

<table>
<thead>
<tr>
<th>Pattern</th>
<th>Pre</th>
<th>Collaborative</th>
<th>Post</th>
<th>Rate (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Correct</td>
<td>Correct</td>
<td>Correct</td>
<td>53</td>
</tr>
<tr>
<td>B</td>
<td>Correct</td>
<td>Correct</td>
<td>Incorrect</td>
<td>1</td>
</tr>
<tr>
<td>C</td>
<td>Correct</td>
<td>Incorrect</td>
<td>Correct</td>
<td>4</td>
</tr>
<tr>
<td>D</td>
<td>Correct</td>
<td>Incorrect</td>
<td>Incorrect</td>
<td>2</td>
</tr>
<tr>
<td>E</td>
<td>Incorrect</td>
<td>Correct</td>
<td>Correct</td>
<td>23</td>
</tr>
<tr>
<td>F</td>
<td>Incorrect</td>
<td>Correct</td>
<td>Incorrect</td>
<td>6</td>
</tr>
<tr>
<td>G</td>
<td>Incorrect</td>
<td>Incorrect</td>
<td>Correct</td>
<td>2</td>
</tr>
<tr>
<td>H</td>
<td>Incorrect</td>
<td>Incorrect</td>
<td>Incorrect</td>
<td>9</td>
</tr>
</tbody>
</table>

5. Conclusions and Future Work

This study proposed learning activities with KB map in a group for knowledge construction and sharing prior to development of creativity, and implemented it to investigate the knowledge propagations among students. The result supports the three hypotheses, “Students share their understandings through discussion,” ”Students’ understandings get close to the correct one” and ”Teachers can teach based on the group map.” Moreover, knowledge propagation in their learning is analyzed from the pre-maps to post- ones by way of collaborative- ones. Consequently, there were every possible patterns and the majority of the propositions changed to correct, however, some propositions changed to wrong.

From these results, students could acquire and correct their knowledge through the proposed group work. In knowledge propagation analyzed in Section 4.2 most of students keep their correct knowledge and correct it through discussion. This tells that the proposed method did not give negative effect on students in this case. KB map gives students with common parts to build a concept map. This might be effective in discussion. If the students built concept maps freely it would be difficult for them to organize their thought at a short time. This is highly controversial issue and must be investigated in the future.

The aim of the classes proposed in this study is to build common knowledge among students before forming the creative opinions in the next class. Therefore, what is important is to let learners be with a full understanding of basic knowledge required for forming creative opinions. To that end, learners must keep the correct knowledge and change the incorrect knowledge to correct. In the proposed classes, the students improved their understanding through group work with KB map after learning the subject. Also, with KB map, the teacher could teach about the propositions that the students did not understand well and conduct personal coaching based on the data from KB map the students made.
These classes were implemented in the regular classes and there is no data to improve the improvement of students’ understanding. It is necessary to measure their understandings with other tests and to compare the effectiveness with learning methods. However, at least, the teacher conduct this classes said that he feels the reciprocal teaching by students have worked well and KB map is useful not only learners but also teachers to organize what to learn and recognize students’ understanding.

Of course, the aim of this study is not general in the research area of collaborative learning. Most of studies focus on discovery learning in which students build their own knowledge other than knowledge given by teachers. However, Scardamalia distinguishes it into "guided discovery" and "knowledge building" (Scardamalia, 2007). The proposed lesson in this study can be considered as a kind of the former one. In this type of discovery learning learners try to discover a solution of a problem that have correct answers. Although, this does not require creativity in the true sense, they are required creative thinking. The authors think that the advantage of this style of discovery learning is the existence of correct answer and teachers can make clear assessment on the answers of learners. This is expected to use for training of creative thinking. In fact, KB map building in the lessons required the students to think about things other than they had learned. They needed to make a conjecture from the materials given in the lessons. Although this is closed in the KB map and the materials, thinking process is similar to inductive thinking in true discovery learning.

In addition to that, further research on this collaborative learning with KB map would clarify the knowledge propagation. A further direction of this study will be to support the classes for acquiring advanced knowledge by collaborative learning with KB map.

References


The Exploration of Improving Efficiency of Synchronous Discussion: e-Case Live Show

I-Fan LIU\textsuperscript{a*}, Chun-Wang WEI\textsuperscript{b}
\textsuperscript{a}Department of Information Management, Oriental Institute of Technology, Taiwan
\textsuperscript{b}Department of Management Information System, Far East University, Taiwan
*ifliug@gmail.com

Abstract: In traditional discussion style, the presentation and discussion of the case is usually led by the team leader, but there are still some members who do not actively participate in the discussion, causing the atmosphere of participation to become disjointed. In order to encourage students actively interactive with instructor and peers, the purpose of the study is to propose an advanced synchronous learning style, e-Case Live Show, which integrates live webcast technique with hosting by turns to support learners’ case-based learning. 58 college students participated in this study. From the research results, we found that students from EG showed significantly higher average scores on Perceived Learning Autonomy and Perceived Interactivity than CG. The e-Case Live Show developed by this study to support case-based learning is indeed better than the traditional TV talk show discussion format for giving learners more autonomy and interactive learning.

Keywords: Case-based learning, synchronous discussion, learning autonomy, learning interactivity

1. Introduction

In the past ten years, Information and Communication Technology (ICT) has developed rapidly, reducing the distance of communication around the world. Learning no longer has to be fixed in time and place, but rather can be flexible. With the age of Web 2.0, social networking services, such as Facebook and YouTube make social interaction and sharing more and more frequent (Hughes, Rowe, Batey, & Lee, 2012). In Taiwan, many professors in business schools conduct the case-based learning in parts of their management courses in order to let students experience the conditions described in cases. Real business operating conditions can be even simulated through various software. One such program, for example, involves students taking a course on production management, in which they can discuss problems faced by enterprises within the context of a simulated supply chain system through case studies, and go through business operations, such as purchase of materials, sale of products, and inventory (Liu & Young, 2006). This software simulation practice will improve students’ learning efficiency. The objective is to go beyond traditional teaching methods of teachers giving their courses to students from a podium.

According to Tian & Hong’s (2003) research, students in Taiwan are less active in asking questions than those in United States and Europe. They also observe the following phenomena in Taiwan’s college students in class of case-base discussion: (1) In traditional teaching method, teachers stand on the podiums to impart knowledge and students sit down to accept it. The problem is that students usually just sit there and do not ask questions or speak up; (2) In traditional discussion style, the presentation and discussion of the case is usually led by the team leader, but there are still some members who do not actively participate in the discussion, causing the atmosphere of participation to become disjointed; (3) Unless in group discussion, students prefer not to interact very often. They usually learn related knowledge
individually and tend to become lone learners; (4) Students can learn and finish their works through group discussion and increase their interaction. However, it is likely that each student only cares about his or her work, integrating with the knowledge from others is difficult. In order to encourage students actively interactive with instructor and peers, the purpose of the study is to propose an advanced synchronous learning style, e-Case Live Show, which integrates live webcast technique with hosting by turns to support learners’ case-based learning. Therefore, we also want to explore whether the efficiency of synchronous discussion style like e-Case Live Show had better than traditional discussion style.

2. The Call-in programs and learning

The TV talk shows originated in the United States in the 1930s from political talk shows and it also known as political call-in programs. The reason why the political call-in program is broadly popular to this day is mainly because it offers citizens and participants in the program with opportunity to engage in two-way communication and express personal opinions (McLeod, Dietram, & Moy, 1999). The call-in programs can raise the participation of discussions, leading to higher ratings. It is a kind of learning when people observe others’ behavior or thought (Horowitz, 1993). Bandura (1986) has proposed observed learning and imitation in a social learning theory. That is, observed learning indicates that an individual, as an onlooker, observes others’ behavior in a process of obtaining learning. Imitation, on the other hand, is the process of learning other individuals' or groups' behavior in observed learning (Chang, 1996). Bandura (1986) argues that the most important concept of social learning theory is vicarious learning, which means people learning and understanding what behaviors are shown in certain conditions by observing others’ behaviors and outcomes.

3. The Case-based Learning Method

The case-based learning method originated from Harvard University in 1890, and has been practiced in the business courses at the Harvard Business School for a long time (Barnes, Christensen, & Hansen, 1994). This method is a unique type of teaching method, in which, as Shulman (1992) points out, educators use cases to familiarize students with the conditions under which cases happen. Through simulating the conditions and discussing possible decisions, students try to develop their capability to solve problems.

The case-based method, considered an effective way to combine theory and practice in business education (Harrington, 1995; Knirk, 1991), focuses on the interaction among peers and the solution from brainstorming (Levin, 1995). There are generally two arrangements in the case-based method, prearranged groups and ad hoc groups (Wassermann, 1994). The former involves a teacher gathering students with different characters in a group, such as talkers and non-talkers. The latter involves students grouping to work with others they are familiar with. Although the case-based method is a self-learning one in which students actively participate in discussion, learners still have to interact and cooperate with others in order to achieve better learning effects (Blumenthal, 1991).

4. e-Case Live Show

The synchronous discussion of e-Case Live Show is based on live webcast technique to conduct a videoconferencing. We adopt i-Share technique, a synchronous teaching module based on
multimedia collaboration system developed by SUNNET Corporation (http://www.sun.net.tw). This module provides multi-player audio and video instant communication like a well-known synchronous discussion tool JoinNet (Chen, Ko, Kinshuk, & Lin, 2005; Chen & Ko, 2010) to offer online face-to-face teaching and learning for instructors and learners (as Figure 1).

Figure 1. e-Case Live Show synchronous learning environment

5. Methodology

In order to evaluate students’ interactive learning efficiency for the e-Case Live Show, the five-point Likert scale and independent sample *t*-test was adopted in this quasi-experiment. The participants were divided into an experiment group (EG) and a control group (CG). Thirty-six students volunteered to enter the EG group; 22 students were in the CG group. For the EG group, each team member had to play the role of leader by turns to host and discuss the content of the case study he/she was responsible for. For the CG group, the format was like a traditional TV talk show in which one leader hosted and discussed the content of the case study with guests. For both EG or CG groups, all of the students enrolled in the class called “Theory of Management and Case Study”, and the professor carried out a blended form of instruction in the course for one semester. The section on management theory (2/3) was taught face-to-face by the professor; the case study (1/3) was carried out through the e-Case Live Show. This was a learner-centered and discussion-based course. Students had to look for partners, and each group consisted of 3 to 4 members. One group was responsible for one case study from the textbook. To prevent the situation where few team members speak or participate, each EG member took turns to be the host during the course of the one-hour online discussion; students in remote areas could also call in to participate in interactive discussions with the host.

6. Findings
From the research results, we found that students from the EG group showed significantly higher average scores on Perceived Learning Autonomy ($M_{EG} = 4.22 > M_{CG} = 3.64$, $t = 2.41$, $Sig = .019 \ast$) and Perceived Interactivity ($M_{EG} = 4.42 > M_{CG} = 3.86$, $t = 2.73$, $Sig = .008 \ast\ast$) than the CG group. This suggests that the students from the EG group experienced much more autonomous and interactive learning in the whole process of case-based learning as compared to the CG group. In terms of Perceived Coordination ($M_{EG} = 3.69 < M_{CG} = 4.23$, $t = -2.01$, $Sig = .049 \ast$), we found that the EG students showed significantly lower average scores than CG ones. This indicates that because students from the EG group had to play a host role by turns, their coordination performance was not as efficient as the CG host. On the other hand, there was no significant difference on Ease of Use ($M_{EG} = 4.53 > M_{CG} = 4.18$, $t = 2.73$, $Sig = .147$) between the two groups.

7. Conclusions

The e-Case Live Show described in this study to support case-based learning clearly outperforms the traditional TV talk show discussion format for giving learners more autonomy and opportunities for interactive learning. Studies in the future could be focused on how to provide support for the students with poor oral communication skills when they are hosting the discussion of a case study, such as by encouraging assistance from their peers to increase their perceived coordination, so that these students can better express themselves.

Acknowledgements

This work was supported by the National Science Council of Taiwan under contract numbers NSC 101-2511-S-269-002-MY3.

References


Game playing as a strategy to improve Team Cohesion, support for collaborative U-Learning

Pei-Yu Cheng\(^a\), Wen-Yen Wang \(^b\) & Yueh-Min Huang\(^*\)
\(^a\) Department of Engineering Science, National Cheng Kung University, Taiwan
\(^b\) Department of Computer Science and Information Engineering, Kun Shan University
\(^*\) huang@mail.ncku.edu.tw

Abstract: This study investigated how a game playing strategy embedded in collaborative u-learning activities affect students’ Team Cohesion and Learning attitude and their learning performance. Participants in this study were fifth grade students in elementary school (N=64); they were randomly assigned into the experiment group and the control group. The Experimental Group and Control Group will be assigned different English activity in the first stage, and in the second stage, two groups conducted the same collaborative u-learning activities. The results indicate that game playing strategy can greatly enhance students’ Team Cohesion and Learning Attitude in our study.

Keywords: Game playing, Team Cohesion, U-Learning, Collaborative

1. Introduction

In Taiwan, English is the important second language. Consequently, enhancing student’s English ability has become important educational policies. For learners, vocabulary knowledge and reading ability are the most important components of performance in second language learning (Huckin, 1995), Folse (2004) indicated that vocabulary is essential to English learning for second-language learners. Therefore, it’s a vital issue to develop a sound approach by which to assist students in learning English vocabulary. In recent years, with the rapid evolution of computer technology and the prevalence of mobile devices, learning has changed transformed from traditional classroom learning to digital and mobile learning. For vocabulary learning, many studies have tried to explore, how to use mobile devices to support vocabulary learning (Chen & Chung, 2008; Hong, Hwang, Tai, & Chen, 2014; Y. M. Huang, Huang, & Lin, 2012).

According to related studies, ubiquitous learning is an effective teaching methods, because combining u-learning can effectively trigger learners’ learning motivation (Chiou, Tseng, Hwang, & Heller, 2010; Jeng, Lu, & Lin, 2010; Ogata & Yano, 2004) and enhance their learning performance (El-Bishouy, Ogata, & Yano, 2007; Rogers et al., 2005). Liu and Chu (2010) indicated that incorporating ubiquitous into the English learning activities could achieve a better learning outcomes and motivation.

According to this viewpoint, many researchers have been interested in ubiquitous learning, and has been successfully applied to many subjects (Y.-M. Huang & Chiu, 2014; Y.-M. Huang, Huang, &
Researchers have pointed out that using mobile devices may enhance collaborative learning and promoted better interactions between students in the activities because students can use it to coordinate collaboration between them. Lai and Wu (2006) argued that using mobile devices can effectively enhance students' attitudes and performance in collaborative learning.

However, an earlier study show that there are many problems of online collaborative learning, such as difficulties in communication, the lack of shared, and the imbalance (Roberts & McInnerney, 2007; Tseng & Yeh, 2013). Therefore, the dynamic within the team is also an important consideration in building Team Cohesion (Kwon, Liu, & Johnson, 2014). Consequently, this study proposed a game playing strategy which is embedded in collaborative u-learning activities for helping students to building Team Cohesion.

2. Research Methods

2.1 Participants

This study investigated how a game playing strategy embedded in collaborative u-learning activities affect students’ Team cohesion and their learning performance. Participants in this study were fifth grade students in elementary school (N=64); they were randomly assigned into the experiment group and the control group.

Figure 1 shows the experimental flow of this study. Before the experiment, this study distributed pretest to subjects to find if there is significant difference between two groups. After the subjects filled out the questionnaire, the researcher conducted experiment on two groups. In the first stage, the students in the experimental group conducted collaborative crossword game as Figure 2. While the students in the control group conducted ordinary learning activities. In second stage, the two groups conducted the same collaborative u-learning activities for learning English vocabulary as Figure 3.

![Figure 1. Experimental procedure](image-url)
3. Result and Discussion

This study used the pre-test scores as covariate for one-way ANOVA to avoid any interaction effects from the pre-test on the students’ learning outcomes. As listed in Table 1, the pre-test mean for the experimental group was 57.97 and 57.81 for the control group. The results did not reach a level of significance, $f=0.005$, $p>.05$. It suggests that homogeneity of two groups of variables is supported.

Table 1: The one-way ANOVA results for the pre-test scores.

<table>
<thead>
<tr>
<th>Group</th>
<th>N</th>
<th>mean</th>
<th>SD</th>
<th>f</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre-test</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Experimental</td>
<td>32</td>
<td>57.97</td>
<td>9.233</td>
<td>0.005</td>
</tr>
<tr>
<td>Control</td>
<td>32</td>
<td>57.81</td>
<td>9.046</td>
<td></td>
</tr>
</tbody>
</table>

According to Table 2, the post-test mean for the experimental group was 72.19, and 70.31 for the control group. Results of statistical analysis showed a no significant difference in learning
performance between two groups, $f=2.416$, $p>.05$. This result suggests that learners in the first stage with different learning activities didn’t produce a significant difference within learning performance.

Table 2: The one-way ANOVA results for the post-test scores

<table>
<thead>
<tr>
<th>Group</th>
<th>N</th>
<th>mean</th>
<th>SD</th>
<th>$f$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Post-test</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Experimental</td>
<td>32</td>
<td>72.19</td>
<td>4.568</td>
<td>2.416</td>
</tr>
<tr>
<td>Control</td>
<td>32</td>
<td>70.31</td>
<td>5.070</td>
<td></td>
</tr>
</tbody>
</table>

* $p < .05$

As listed in Table 3, the experimental group students' Team Cohesion were significantly higher than the control group students, $t=8.99$, $p<.001$). That is, the students who conducted game playing strategy had higher Team Cohesion than those who conducted ordinary learning activities in the first stage.

Table 3: The one-way ANOVA results of Team cohesion

<table>
<thead>
<tr>
<th>Group</th>
<th>N</th>
<th>Mean</th>
<th>SD</th>
<th>$f$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Team Cohesion</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Experimental</td>
<td>32</td>
<td>4.36</td>
<td>0.38</td>
<td>8.99**</td>
</tr>
<tr>
<td>Control</td>
<td>32</td>
<td>4.12</td>
<td>0.23</td>
<td></td>
</tr>
</tbody>
</table>

*p < .05

As listed in Table 4, the experimental group students' Learning Attitude were significantly higher than the control group students, $t=-17.384$, $p<.001$). That is, the students who conducted game playing strategy had higher Learning Attitude than those who conducted ordinary learning activities in the first stage.

Table 4: The one-way ANOVA results of Learning Attitude

<table>
<thead>
<tr>
<th>Group</th>
<th>N</th>
<th>Mean</th>
<th>SD</th>
<th>$f$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Learning Attitude</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Experimental</td>
<td>32</td>
<td>4.33</td>
<td>0.23</td>
<td>17.384***</td>
</tr>
<tr>
<td>Control</td>
<td>32</td>
<td>3.95</td>
<td>0.46</td>
<td></td>
</tr>
</tbody>
</table>

*p < .05

4. Conclusion

In this study, we proposed a game playing strategy which is embedded in collaborative u-learning activities for helping students to building Team Cohesion. Based on the experimental results, we found that game playing strategy can greatly enhance students’ Team Cohesion and Learning Attitude which is consistent with the findings of the past research (DeVries & Edwards, 1973; Huyen & Nga, 2003; Randel, Morris, Wetzel, & Whitehill, 1992; Roberts & McInerney, 2007; Sharan & Sharan,
Thus, we suggest that teachers can use game playing as a strategy to improve Team Cohesion, support for collaborative U-Learning. This study has certain limitations, such as manpower. The limitation of this study is too small of sample size in this experiment. In the future research, we will consider some experiments with a larger sample size of students and conduct more complete research to investigate the relationships between Team Cohesion and Learning Styles.

Acknowledgements
This study is conducted under the "Cloud computing systems and software development project (3/3)" of the Institute for Information Industry which is subsidized by the Ministry of Economy Affairs of the Republic of China.

References


Computer Supported Collaborative Learning: Comparison between good vs. poor collaborators.

Computers & Education.


