Lap-Kei Lee · Leong Hou U ·
Fu Lee Wang · Simon K. S. Cheung ·
Oliver Au · Kam Cheong Li (Eds.)

Communications in Computer and Information Science 1302

Technology in Education

Innovations for Online Teaching and Learning

5th International Conference, ICTE 2020
Macau, China, August 19–22, 2020
Revised Selected Papers

Springer

Communications in Computer and Information Science 1302

More information about this series at http://www.springer.com/series/7899

Lap-Kei Lee · Leong Hou U ·
Fu Lee Wang · Simon K. S. Cheung ·
Oliver Au · Kam Cheong Li (Eds.)

Technology in Education

Innovations for Online Teaching and Learning

5th International Conference, ICTE 2020
Macau, China, August 19–22, 2020
Revised Selected Papers

Springer

Editors
Lap-Kei Lee
The Open University of Hong Kong
Hong Kong, China

Leong Hou U
University of Macau
Macau, China

Fu Lee Wang
The Open University of Hong Kong
Hong Kong, China

Simon K. S. Cheung
The Open University of Hong Kong
Hong Kong, China

Oliver Au
The Open University of Hong Kong
Hong Kong, China

Kam Cheong Li
The Open University of Hong Kong
Hong Kong, China

ISSN 1865-0929 ISSN 1865-0937 (electronic)
Communications in Computer and Information Science
ISBN 978-981-33-4593-5 ISBN 978-981-33-4594-2 (eBook)
https://doi.org/10.1007/978-981-33-4594-2

This Springer imprint is published by the registered company Springer Nature Singapore Pte Ltd.
The registered company address is: 152 Beach Road, #21-01/04 Gateway East, Singapore 189721, Singapore

Preface

This edited volume consists of the extended papers selected from the 5th International Conference on Technology in Education (ICTE 2020), which was hosted by University of Macau in Macao, China during August 19–22, 2020, and was held online due to the COVID-19 pandemic.

Technology has become an integral part in virtually all aspects of education, broadly covering curriculum planning, content development and delivery, communication among learners, instructors and institution, assessment and program evaluation. Enabled by the latest technological advances, new and innovative measures are derived to improve teaching and learning effectiveness. Taking "Innovations for Online Teaching and Learning" as its main theme, ICTE 2020 provides a platform for knowledge exchange and experience, sharing among researchers and practitioners in the field of technology in education.

ICTE 2020 attracted a total of 79 submissions. After a careful paper review process, 30 papers were selected for inclusion in this volume. These papers are organized in five groups: (1) instructional technology, (2) learning analysis and assessment, (3) learning environment, (4) open and collaborative learning, and (5) technology and education.

Our sincere thanks go to the conference's Organizing Committee for their effective administration and unfailing support. Our thanks also go to the International Program Committee. The high quality of the papers could not have been maintained without their professional comments and advice in the paper-review process.

October 2020

Lap-Kei Lee
Leong Hou U
Fu Lee Wang
Simon K. S. Cheung
Oliver Au
Kam Cheong Li

Organization

Organizing Committee

Honorary Chair

John MacIntyre	University of Sunderland, UK

Conference Chairs

Oliver Au	International Hybrid Learning Society, UK
Kam-Weng Tam	University of Macau, Macau

Programme Chairs

Leong Hou U	University of Macau, Macau
Lap-Kei Lee	The Open University of Hong Kong, Hong Kong

Organizing Chair

Emily Pey-Tee Oon	University of Macau, Macau

Financial Chairs

Kwan-Keung Ng	University of Sunderland, UK
Miguel Gomes da Costa Junior	University of Macau, Macau

Registration Chairs

Jeanne Lam	University of Hong Kong, Hong Kong
Johnny Keng Weng Lao	University of Macau, Macau

Members

Simon K. S. Cheung	The Open University of Hong Kong, Hong Kong
Wai Shing Ho	Caritas Institute of Higher Education, Hong Kong
Ivan Ka Wai Lai	City University of Macau, Macau
Kam Cheong Li	The Open University of Hong Kong, Hong Kong
Will Ma	Technological and Higher Education Institute of Hong Kong, Hong Kong
Chung Keung Poon	The Hang Seng University of Hong Kong, Hong Kong
Fu Lee Wang	The Open University of Hong Kong, Hong Kong
Leung-Pun Wong	Tung Wah College, Hong Kong

International Program Committee

Chairs

Leong Hou U	University of Macau, Macau
Lap-Kei Lee	The Open University of Hong Kong, Hong Kong

Members

Oliver Au	The Open University of Hong Kong, Hong Kong
Gary Cheng	The Education University of Hong Kong, Hong Kong
Simon K. S. Cheung	The Open University of Hong Kong, Hong Kong
Wing Sum Cheung	Nanyang Technological University, Singapore
John Kwok Tai Chui	The Open University of Hong Kong, Hong Kong
Hew Khe Foon	The University of Hong Kong, Hong Kong
Miguel Gomes da Costa Junior	University of Macau, Macau
Wai Shing Ho	Caritas Institute of Higher Education, Hong Kong
John Hui	City University of Hong Kong, Hong Kong
Gwo-Jen Hwang	National Taiwan University of Science and Technology, Taiwan
Ivan Ka Wai Lai	City University of Macau, Macau
Kam Cheong Li	The Open University of Hong Kong, Hong Kong
Will Ma	Technological and Higher Education Institute of Hong Kong, Hong Kong
Kwan-Keung Ng	University of Sunderland, UK
Shaun Nykvist	Queensland University of Technology, Australia
Emily Pey-Tee Oon	University of Macau, Macau
Chung Keung Poon	The Hang Seng University of Hong Kong, Hong Kong
Kam-Weng Tam	University of Macau, Macau
Jeff K. T. Tang	Hong Kong Productivity Council, Hong Kong
Fu Lee Wang	The Open University of Hong Kong, Hong Kong
Chun-Wang Wei	Kaohsiung Medical University, Taiwan
Tak-Lam Wong	Douglas College, Canada
Yau-Yuen Yeung	The Education University of Hong Kong, Hong Kong
Y. T. Yu	City University of Hong Kong, Hong Kong
Zehui Zhan	South China Normal University, China

Organizer

INTERNATIONAL
HYBRID LEARNING SOCIETY
LEARN EVERYWHERE, EVERY WAY

International Hybrid Learning Society

Co-organizers

University of Macau

Tung Wah College

University of Sunderland

Sponsor

Hong Kong Pei Hua Education Foundation

Contents

Instructional Technology

A Case Study on C-STEAM Education: Investigating the Effects of Students' STEAM Literacy and Cultural Inheritance Literacy 3
Liming Huo, Zehui Zhan, Ziying Mai, Xiao Yao, and Yunxiang Zheng

The Experiential Learning Unit for Promoting Students' Understanding of Vapor Pressure and Related Concepts 13
Siwa Mahardthai, Supan Yodyingyong, Namkang Sriwattanarothai, and Piyachat Jittam

A Corpus-Based Analysis of Text Complexity in NMET Reading and Its Implication on Reading Instruction 23
Liyan Huang, Jiaying Wang, and Xiaochen Hua

A Mobile Game-Based Learning Approach for Motivating Preschoolers and Primary Students in Learning Mathematics. 35
Chok-pang Kwok, Sin-chun Ng, Ching-hei Ho, So-sum Ip, and Chi-wang Daniel Kui

Evaluating Online Tourism Education Quality by Using an Instructional Strategy Framework 46
Xialei Duan, Chung-Shing Chan, and IpKin Anthony Wong

The Impacts of Digital Note-Taking on Classroom Instruction: A Literature Review 61
Harrison Hao Yang, Yinghui Shi, Huiyun Yang, and Qiuyu Pu

Learning Analysis and Assessment

An Empirical Study of the Effect of ASR-Supported English Reading Aloud Practices on Pronunciation Accuracy 75
Yuting Yuan and Xiaobin Liu

Improving English Phoneme Pronunciation with Automatic Speech Recognition Using Voice Chatbot 88
Victor Chi-Wa Cheng, Vincent King-Tang Lau, Ringo Wai-Kit Lam, Tian-Jie Zhan, and Pak-Kin Chan

The Design and Implementation of a Computerized Adaptive Testing System for School Mathematics Based on Item Response Theory 100
Jiyou Jia and Huixiao Le

Medical Teachers' Action Research: The Application of Bloom's Taxonomy in Formative Assessments Based on Rain Classroom 112
Wei Li, Hongmei Cai, Xuesong Yang, and Xin Cheng

Automatic Generation of Matching Rules for Programming Exercise Assessment 126
Tak-Lam Wong, Chung Keung Poon, Chung Man Tang, Yuen Tak Yu, and Victor Chung Sing Lee

Automatic Question Generation System for English Reading Comprehension 136
Yin-Chun Fung, Jason Chun-Wai Kwok, Lap-Kei Lee, Kwok Tai Chui, and Leong Hou U

Learning Environment

Social Media Acceptance Scheme 149
Miloslava Cerna

Exploring the Effect of Multichannel Multimodal Learning Environment on Student Motivation and Self-efficacy. 164
Jun Xiao, Tzu-Han Lin, and Hong-Zheng Sun-Lin

Establishing a Tripartite Intelligent Reading Platform Connecting Schools, Families, and Students: An Approach with Chinese Characteristics to Promote Students' Reading. 176
Jielin Zeng, Yirao Liu, Xiaoxiao Huang, Ruixiang Gao, Hu Mei, and Zehui Zhan

A Study on College Students' Self-regulated Online Learning in the Home Study Context 189
Wenyi Chen

Methods and Techniques to Improve International Learning Experiences Online 201
Wenbin Ji, Pan Li, Meihua Wang, and Jinbin Zhang

Improving Productive Classroom Talk Through Visual Learning Analytics Technology: A Case Study of an Award-Winning Mathematics Teacher 213
Chung Kwan Lo and Gaowei Chen

Open and Collaborative Learning

Supporting Social Regulation of Learning with Multi-dimensional Scaffolding in Computer-Supported Collaborative Writing Activities 227
Kailiang Chen, Haogang Bao, Yanyan Li, You Su, and Yu Peng

Facilitating Course Recommendations by Word2vec Paradigm Through Social Tags 239
Jingjing Wang, Haoran Xie, Oliver Tat Sheung Au, Lap-Kei Lee, Di Zou, and Fu Lee Wang

Social Media in Higher Education: A Review of Their Uses, Benefits and Limitations 248
Kam Cheong Li and Billy Tak-Ming Wong

The Impact of AWE and Peer Feedback on Chinese EFL Learners' English Writing Performance 258
Yanna Xie, Liyan Huang, and Yilu Wang

A Review of Small Private Online Courses in Higher Vocational Colleges . . . 271
Ning-wei Sun and Simon K. S. Cheung

Enhancing Academic Reading Skills Using a Peer Assessment of Online Collaborative Annotation Approach 281
Wing Shui Ng, Haoran Xie, and Gary Cheng

Technology and Education

Influence Evaluation of Academic Papers via Citation Characteristics Analysis 295
Qingyuan Wu, Jiahua Hu, Yanghui Rao, Fu Lee Wang, and Haoran Xie

Users' Behavioral Intention Toward M-Learning in Tourism English Education: A Case Study of Macao 308
Xiaohong Wu, Chi Man Tam, and Sha Fang

Leveraging Neural Network-Based Model for Context Classification of Classroom Dialogue Text 323
Shunwei Lei, Jinlin Li, Yu Song, Yingshan Shen, Lap-Kei Lee, and Tianyong Hao

Application of Blockchain in Academic Credit Bank System 337
Ning Yan

Using Education Technologies to Accommodate Vocational and Professional Education Training (VPET) Students' Attributes 349
Kwan-Keung Ng, Ricky Yuk-Kwan Ng, and Rechell Yee-Shun Lam

Recognition of Roles of Variables Based on Deep Learning Technologies . . . 362
Andrew Kwok-Fai Lui, Tsz-Tik Lui, Tommy Cheuk-Hin Leung, and Ho-Yin Chan

Author Index 375

Instructional Technology

A Case Study on C-STEAM Education: Investigating the Effects of Students' STEAM Literacy and Cultural Inheritance Literacy

Liming Huo, Zehui Zhan(✉), Ziying Mai, Xiao Yao, and Yunxiang Zheng

School of Information Technology in Education, South China Normal University, NO. 55 Zhongshan Road West, Guangzhou 510631, China
zhanzehui@m.scnu.edu.cn

Abstract. C-STEAM is a localized transdisciplinary education in Chinese context with the goals of inheriting traditional culture and cultivating students' STEAM literacy. It mainly has three core values: educational value of cultivating students' core literacy; carrier value of inheriting outstanding traditional culture; and social value of booming regional culture. In order to see how a C-STEAM project works and effects, we adopted "6C" Instructional Design Model which is an operational steps including Contextual Experience, Connotation construction, Characteristic inquiry, Create artifact, Connect with society and Conclusion reflection. In a case study lasting 6 weeks on teaching Cantonese slang in a primary school in Foshan, China, questionnaires were used to analyze the students' learning performance, STEAM literacy, cultural understanding and inheritance, and structured interviews were used to collect views and reflections of teachers and students. Results indicated that the application of "6C" Model has significantly improved students' cultural understanding and cultural identity and cultural inheritance, and shown that C-STEAM plays an important role as the cultural carriers, brings some social value for constructing regional characteristic culture and enables to cultivate students' core literacy and improves students' transdisciplinary learning ability. More research is needed to ameliorate "6C" Model to achieve the core values of C-STEAM.

Keywords: C-STEAM education · "6C" instructional design model · Core values · STEAM literacy

1 Introduction

Chinese nation has a large number of splendid cultural treasures in the past five thousand years, which is worth inheriting and developing by every generation. However, due to the development and evolution of social economy, the invasion of foreign cultures, the lack of cultural inheritors and protectors and other reasons, the rich traditional Chinese culture has been facing the problems of being destroyed, and even disappeared (Zhang 2014). Therefore, it is urgent to inherit and innovate the traditional Chinese culture. Meanwhile, Chinese traditional culture has been ignored in elementary education. Therefore,

L.-K. Lee et al. (Eds.): ICTE 2020, CCIS 1302, pp. 3–12, 2020.
https://doi.org/10.1007/978-981-33-4594-2_1

educators (Li 2018) have the responsibility to introduce Chinese traditional culture into the schooling system (Zhou 2016; Guo and Jin 2018). Booming of STEAM education has enabled chances for traditional culture to enter elementary education (Shou and Hu 2017) and provided students with opportunities for in-depth exploration and cultural innovation (Zhao 2017).

Therefore, the localized STEAM education proposed in this article (short for "C-STEAM") is an interdisciplinary education aiming at inheriting traditional Chinese culture; that is, promoting traditional Chinese culture and national emotions (Zhan et al. 2020). "C" in C-STEAM education that is short for culture, is a target-oriented concept, that is, to take cultural inheritance as the main educational goal.

We confirm the core values of C-STEAM, and try to find out the implementation path and effect of the core values in a case study which follows "6C" instructional design model that we adopted, and serves as a conductive framework to carry out the C-STEAM project. And it will enrich research and practice of C-STEAM and offer new possibilities for the localization of STEAM education in China.

2 The "6C" Instructional Design Model

At present, a few schools in China have tried to explore C-STEAM on the practical level, and it is generally based on the background of traditional Chinese culture at the level of discipline knowledge integration that, in turn, resulted in lacking of systematism, excavation of traditional cultural connotation at the level of values internalization, as well as the analysis of C-STEAM teaching cases and the summary of instructional model. On this basis, we try to apply 6C model, and verify the effectiveness of the model and the influence of C-STEAM on the core values in a case study.

The 6C model was formed by combining three models (Zhan et al. 2020): (1) Interdisciplinary learning activity 5EX design model (Li and Li 2019). (2) Reverse engineering teaching model (Kang and Zhong 2019). (3) Creativity-innovation Thinking interconnection Model (CITIM) (Zhan et al. 2019) Considering the target orientation of cultural inheritance, this study combines the characteristics of C-STEAM, and it summarizes it into six steps, namely Context Setup (C1), Connotation Construction (C2), Characteristic Inquiry (C3), Create Artifact (C4), Connect with Society (C5), and Conclusion Reflection (C6), to create a C-STEAM model for cultural inheritance, as shown in Fig. 1. The first letter of each step is C, so we name the model as 6C model for short.

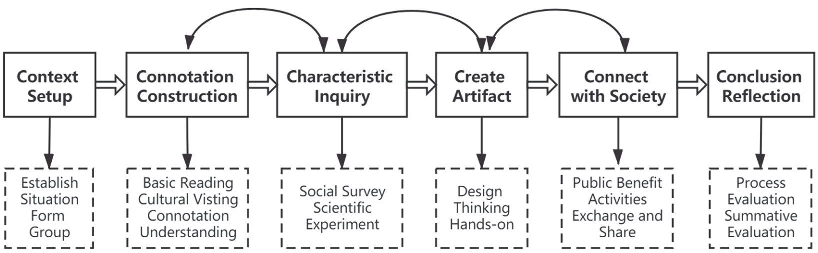


Fig. 1. The "6C" instructional design model for C-STEAM

3 Method

To verify the effectiveness of 6C model, we conducted a case study in a STEAM course for teaching Cantonese slang in a primary school in Foshan, China, and analyzed it based on the data collected from questionnaire, interviews and observation. The cultural theme of the project was Cantonese slang, which is a kind of easy to understand language with local characteristics summed up by Cantonese people in their daily life, bearing unique culture of Cantonese.

3.1 Participants

This case study was conducted in a primary school in Foshan, China. Participants are 30 students (16 males and 14 females) selected from fifth-grade. Among them, 18 were Cantonese speakers, and the remaining 14 were not locals. Students can inquiry independently and learning in a cooperative way, showing strong interest in Foshan local culture, so they are suitable to participate this project with traditional culture as the theme. Before this experiment, participants haven't learned Cantonese slang formally and systematically, only have contacted with some locals in their daily life.

3.2 The Design of the Case

(1) Learning Objectives. The case concentrates on excavating cultural connotation and characteristics of Cantonese slang. Specifically, the teaching objectives follows: 1) Know the basic knowledge of Cantonese (i.e., the historical development, the corpus or language materials); 2) Explore and understand the traits and connotation of Cantonese slang culture; 3) Select appropriate themes and forms based on the characteristics and connotation of Cantonese slang, and then cooperatively design and produce a cultural artifact- Cantonese slang desk calendar; 4) Learn how to organize public service activities and attend to exhibition or visiting; 5) Learn to take self-reflection and evaluation through teamwork.

(2) The "6C" implementation. The project lasting 12 lectures in 6 weeks, below we will discuss the implementation of each step of 6C model briefly.

1) Context Setup (C1). It is composed of four learning activities: creating a situation, asking questions, forming a group and clarifying tasks. First of all, students create situations by reading news of the disappearance of Cantonese and watching the street interview video of Cantonese slang. Secondly, teacher asked questions related to the topic. For example, after watching the video, students were asked if you know the origin of slang and the story behind it? Then, students are divided into several groups, and then defined course's theme and task.

2) Connotation Construction (C2). This part consists of four activities: basic reading, cultural observation, skill training and connotation understanding, which enables students to construct the cultural connotation of Cantonese slang. In basic reading, students read the history materials of Cantonese slang to understand its origin and development of Cantonese slang, in order to cultivate students' STEAM reading and writing ability. In cultural observation part, students watching several video clips of Cantonese slang, such as the story "avalanche barks dog" and other vivid animation videos. In skill training part, teacher gives many examples of using Cantonese slang in daily life, so that students can understand and speak and use them properly in daily life. In the part of connotation understanding, students watch the video clips and perceive the cultural connotation.

3) Characteristic Inquiry (C3). Students mainly inquiry the Cantonese slang cultural connotation through data collection and social investigation. First, students collect Cantonese slang sentences and stories, and then classify them. Secondly, according to the cultural characteristics of Cantonese, students go to local community to make a survey on local residents and masters. Finally, they need to screen and analyze the collected date the authentic cultural characteristics of Cantonese slang were discussed, and summarized the cultural connotations.

4) Create Artifacts (C4). It consists of design and hands-on activities. Firstly, students need to brainstorm to think about what kind of cultural and creative products can be created. In this case, they decided to make a Cantonese slang calendar. Then, students chosen themes that can reflect the cultural features. Then sketches were drawn on the working papers, as well as design intention and illuminations. In the design process, we should actively communicate and absorb various opinions for adjustment. Besides, students should integrate hand-made technology and art drawing into calendar production. Lately, students were suggested to choose production materials and complete the preliminary products. The produce processes include recording and analyzing problems, proposing solutions, collecting suggestions form classmates and teachers, and finally accomplished their products making.

5) Connect Society (C5). It allows students to exchange, share and display their achievements through public welfare activities, exhibitions and artistic performances. In the beginning, students complete the preliminary preparation work of the public activities, including the design of the activity plan, poster design, the rehearsal of performances, etc. Then confirm the activity site, invite local residents, teachers, friends and parents to participate, and carry out activities in some forms such as charity sale and Cantonese slang performances. Finally, students record the activities, feelings, and share thoughts with each other.

6) Conclusion Refection (C6). It uses many evaluation methods to evaluate students' learning attitude and effect, like process evaluation and summary assessment, personal reflection and group mutual evaluation, including the test of basic knowledge, the evaluation of stage learning task completion, PMIQ self-reflection table. In the stage learning task, students evaluate task completion and performance in both individual and teamwork way. After course ending, students are suggested to finish a test of basic knowledge and the PMIQ self-reflection form that mainly include the knowledge of the cultural characteristics of Cantonese slang, the knowledge of the cultural connotation, the knowledge of desk calendar making, etc., to examine their gains.

Besides, in the whole process of C-STEAM project, three scaffolding tools were provided to students to get involved fully in learning activities: Guidelines; Learning record document; Assessment sheets.

3.3 Measure Instruments

We collected data with two kinds of questionnaires, namely cultural understanding and inheritance literacy level questionnaire, STEAM literacy level questionnaire. Both of them were used to verify the educational value of 6C model for the educational value of C-STEAM in cultivating students' core literacy. The completion of students' works and the course evaluation are used to check out the carrier value of C-STEAM in inheriting outstanding traditional culture. And the teacher-student interview was used to demonstrate the social value of C-STEAM in creating regional characteristic culture. After that, data were analyzed from two aspects.

(1) A survey of STEAM literacy. The STEAM literacy questionnaire is adapted from Cai (2017) to measure the basic level of STEM of experimental subjects from three dimensions (i.e., STEAM ability, STEAM attitude and STEAM thinking). And the dimension of STEAM attitude that has 15 questions and coefficient of .92 alpha was referenced to Mahoney (2010) and compiled with Likert scale of five. It is composed of three dimensions and has a total of 75 points. The higher the score is, the more consistent the students are with the content description.
(2) A survey of cultural understanding and inheritance literacy level. " Culture Competency " is regarded as the core of all literacy and behavior, and it provides guidance for Critical Thinking, Creativity, Communication and Collaboration as the goal value orientation. And it can be divided into three elements, namely cultural understanding, cultural identity and cultural practice which are the basis for the questionnaire of this study. The culture identity items were adapted from Phinney and Ong (2007) of which alpha was .81 and reliability was .74 and compiled with Likert scale of five. It was made up with a total of 15 items within three dimensions, and the total score was 75 points.

4 Results

4.1 Analysis of STEAM Literacy

Both before and after the C-STEAM project, we use the STEAM literacy scale as post-test and pre-test measurement to test the subjects twice, and analyze the reliability of the scale by SPSS. Results demonstrate that Cronbach's alpha was 0.914, more than 0.9, indicating that the scale has good reliability and stability. Then paired sample T-test was carried out on the pre-test and post test data of each dimension of STEAM literacy. Results show that statistical significance of pre-test and post-test in these three dimensions is 0.00, less than 0.05, and average score of post-test is higher than that of pre-test that means teaching effect is very significant for students' STEAM ability, STEAM thinking and STEAM attitude. And we found that the means of post-test were higher than that of pre-test after comparison of means.

(1) STEAM Ability. The mean value of post test data is higher than that of pre-test. Among them, item T1 "I can accurately identify the problem from the given problem situation, and the basic requirements and limiting conditions to solve the problem" has a small difference (0.56); while item T5 and item T8 have a significant difference. The data of the post test is significantly higher than of the pre-test, which shows that the students' ability of identify problems, to operate and to summarize and reflect has been improved. After the interview, it is found that students think more about the problem to be solved when designing calendar, and spend more time and energy into the production. Therefore, the ability activities in these two aspects have been greatly improved. As for the ability of summary and reflection, after the interview, we learned that after "communication and sharing" and course evaluation activities, they think that their summary and reflection ability has been improved. FromT1 to T9 before and after the test, STEAM capacity dimension is significantly different.

(2) STEAM Thinking. The means of post-test of each item is higher than that of the pre-test. The average value of the post test data of item T11 "I can systematically consider the influence of various factors on problem solving, creatively solve problems", item T12 "when facing more complex problems, I can think of using certain procedures or steps to solve problems" is significantly improved, which shows the comprehensive thinking ability of students, the ability to solve problems step by step is strengthened. The main reason is that students' competencies have been fostered during scheme and artifacts design activities. Item T9 has a relatively small difference, which is 0.6, indicating that students' thinking ability of solving problems by comprehensively using various disciplines has not been significantly improved. There are two main reasons. One is that means of students' pre-test data is relatively high, which is 3.9. The other is that the cultural product of this case is desk calendar, involving relatively little knowledge of various disciplines. However, by comparing mean of T9 to T10, it is found that there is still significant difference in STEAM thinking dimension.

(3) STEAM Attitude. The post-test mean of items T13 to T15 is higher than that of pre-test, and the mean difference of item T13 "I am very interested in science, technology, engineering, humanities and mathematics knowledge" is the most significant, which shows that students' interest in STEAM knowledge has been increased after this course, so there was significant difference in STEAM attitude dimension.

After comparing and analyzing of the mean values of pre-test and post-test data of each item in the STEAM literacy questionnaire (as shown in Fig. 2), it was found that the means of post-test were higher than that of pre-test.

4.2 Analysis of Cultural Understanding and Inheritance Literacy Level

We developed a cultural understanding and cultural inheritance literacy scale which is composed of three dimensions and has been tested twice (i.e., pre-test and post-test), to investigate the changes of students' cultural understanding and cultural inheritance literacy after the Cantonese slang project. A total of 30 questionnaires were issued, with a recovery rate of 100%. The reliability of Cronbach's alpha was 0.916, greater than 0.9, indicating that the scale has good reliability and stability. The results of paired sample T-test show that the post-test of cultural understanding and cultural inheritance literacy is higher than the mean of the pre-test, among which mean value of cultural understanding is the most obvious, indicating that students have a deeper understanding of Cantonese culture after project. In addition, the mean of the post-test of cultural identity was as high as 4.81, indicating that students' sense of identity with traditional culture had been

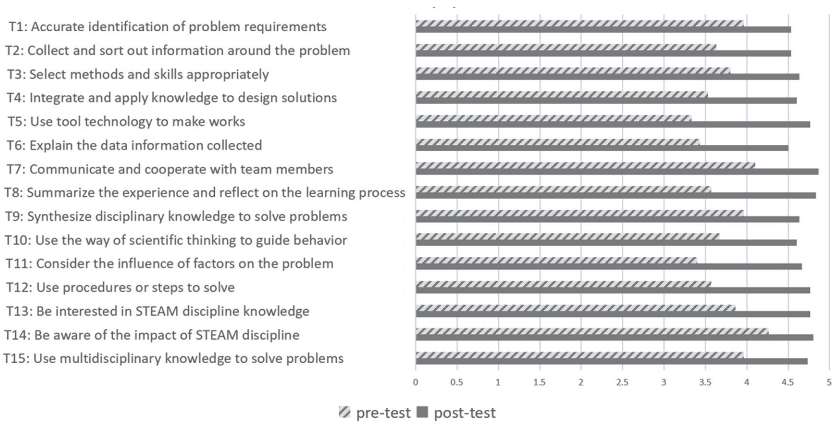


Fig. 2. Comparison and analysis of mean values of pre-test and post-test data of each item in the STEAM literacy questionnaire

greatly improved. The significant probability of three dimensions of the results of pre-test and post-test is 0.00, both less than 0.05, indicating that there is a significant difference in students' cultural understanding and cultural inheritance accomplishment. We also found that the mean values of post-test were higher than that of pre-test (as shown in Fig. 3).

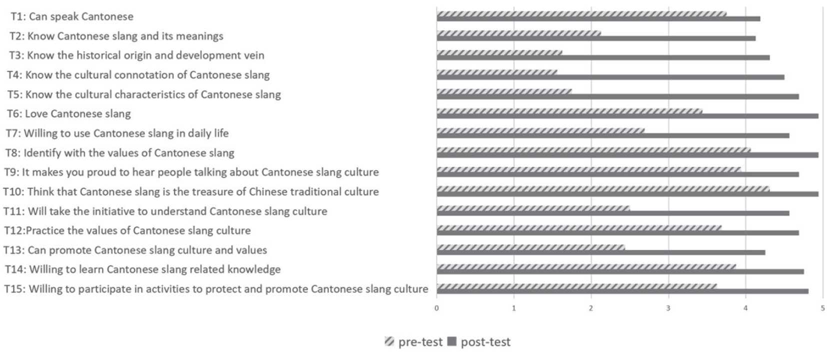


Fig. 3. Cultural understanding and inheritance literacy pre-test and post-test each item mean analysis

(1) Cultural understanding. The means of post-test of item T3 (I know Cantonese historical origin and development), T4 (I know cultural connotation of Cantonese slang), T5 (I know the cultural features of Cantonese slang) were significantly increased. It shows that students have a deeper understanding of Cantonese history, cultural connotation and characteristics after course. The main reason is students have never systematically learned the related knowledge of Cantonese slang before, after a series of activities including basic reading, culture visiting, skill training, and other kinds of activities, students can improve their understanding of Cantonese slang. From the comparison of pre-test and post-test mean values of questions T1 to T5, it can be concluded that there are significant differences in cultural understanding level.

(2) Cultural identity. The mean of post-test for each item is higher than that of pre-test. For example, the mean of the post-test for items T6 (I like Cantonese slang) and T7 (I would like to use Cantonese slang in my daily life) was significantly improved. This indicates that students' love for Cantonese slang and their intention to use it in daily life have been improved. After the interview, it is found that the reason is that students have learned more interesting and common Cantonese slang during the course. They think that Cantonese slang is interesting and can be used in some daily occasions. Through comparison mean values of items T6 to T10, we found that there are significant differences at the level of cultural identity.

(3) Cultural practice. Item T13 to T15 average are higher than the previous test after test, item T11 (I will take the initiative to go to understand Cantonese slang culture), item T13 (I will actively publicize Cantonese slang culture and its values to people around me in my daily life), item T15 (if there is a cultural activities designed to protect and promote Cantonese slang, I am very happy to attend) have the most significant mean difference. This shows that students' liking degree of Cantonese slang cultural and willingness to use it have improved. The reason is that the students found that Cantonese slang exists the phenomenon of cultural losing, through a series of activities, students' sense of responsibility to inherit and promote Cantonese culture has been enhanced which indicates that there is a significant difference.

This paper constructs 6C model, and summarized it into six steps: Context Setup, Connotation Construction, Characteristic Inquiry, Create Artifact, Connect with Society and Conclusion Reflection. In order to confirm the effectiveness of 6C model, we conducted a case study for teaching Cantonese slang in a primary school in Foshan, China. The variables of 6C model were tested by paired sample T-test, the correlation analysis of teaching effect and the core value of C-STEAM project then verify the core values of C-STEAM (as shown in Fig. 4).

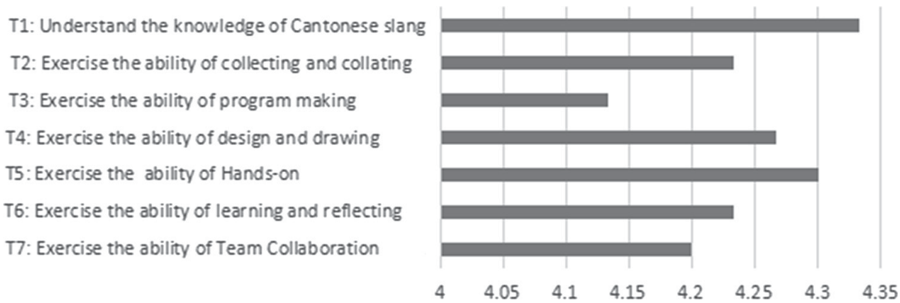


Fig. 4. The influence of Cantonese slang STEAM curriculum on students' literacy ability

5 Discussion and Conclusion

5.1 The Educational Value in Cultivating Students' Core Literacy

C-STEAM education, namely the transdisciplinary education oriented to cultural inheritance, has a strong learning and exploration feature of cultural foundation. The data analysis of learners' cultural understanding and identity shows that the post-test data are significantly higher than pre-test, which proves to some extent that 6C model is conducive to strengthening students' cultural foundation, among which cultural understanding is the most significant. After Cantonese slang project, students have a deeper understanding of Cantonese slang culture, and their sense of identity with traditional culture has been greatly improved, as well as students' cultural practice, STEAM ability, STEAM thinking and STEAM attitude. It indicated that students' cultural foundation, independent development and social participation have been effectively cultivated and improved which proves that C-STEAM education has the educational value of cultivating students' core literacy.

5.2 The Carrier Value of Inheriting Outstanding Traditional Culture

C-STEAM takes Cantonese culture into culture courses to become a carrier of slang dissemination. Students have more opportunities to learn traditional culture deeply on the culture course, which also increases their enthusiasm for learning. In the process of C-STEAM, problem-based and project-based learning are adopted to encourage students to "learn by problems, doing and playing" and that result in "activities" of the traditional culture. The cultural products formed make Cantonese culture a visible, audible and touchable physical product, which makes the traditional culture in a live state. From this point of view, C-STEAM education has the carrier value of inheriting traditional culture.

5.3 The Social Value of Creating Regional Characteristic Culture

Offering C-STEAM courses, schools can formulate local characteristics culture curriculum and then conducive to the construction of specialist schools, featured courses and highlight the educational value of culture, and also promote the transmission and inheritance of traditional culture in the youngsters. This case focus on exploring the connotation and characteristics of Cantonese slang culture, so that students can design and make cultural products with desk calendar as the carrier. In the whole process, students have integrated Cantonese knowledge to achieve the purpose of effectively inheriting and spreading Cantonese slang culture, which better reflects the social value of C-STEAM education in building regional characteristic culture. When students participate in learning activities, they can bring certain communication and promotion effects to regional culture through various channels and methods, and bring corresponding social and economic values. Therefore, C-STEAM curricula not only contribute to the formation of local cultural courses and play the value of cultural education, but also have a positive impact on students' awareness of cultural inheritance and innovation. From this perspective, it is verified that C-STEAM education has the social value of building regional characteristic culture.

Above all, 6C model plays a significant role in promoting traditional culture, not only cultivates students' core qualities such as transdisciplinary thinking and exploration ability, help students to understand deeply cultural connotations and characteristics, but also helps to cultivate and develop the student's humanistic spirit, strengthen the cultural identity and understanding and national self-confidence, so to accomplish the three core values of C-STEAM. In the future, we will enlarge study of 6C model with different cultural themes, to further improve its universality, and build more instructive teaching strategies and tools to enrich the theory and practice of C-STEAM education.

Acknowledgements. This study was supported by the Major basic research and applied research projects of Guangdong Education Department (#2017WZDXM004), MOE (Ministry of Education in China) Project of Humanities and Social Sciences (#19YJC880125); National Social Science Foundation of China (#18BGL053). We also want to thank all the participants in this study.

References

Cai, H.: The study of design and practice of STEM instructional model. Master's thesis, East China Normal University (2017)

Guo, Z., Jin, S.: Excellent traditional Chinese culture. The development and practice of Bo Wu curriculum. J. Chin. Soc. Educ. (S2), 14–21 (2018)

Kang, S., Zhong, B.: Construction of reverse engineering instruction model in robot education. Mod. Distance Educ. Res. **31**(4), 56–64 (2019)

Li, K., Li, Y.: 5EX design model for interdisciplinary learning activities in STEM education. E-Educ. Res. **40**(4), 5–13 (2019)

Li, L.: The coexistence education model of "STEM", "PBL" and "traditional culture". Educ. Sci. Forum (32), 27–30 (2018)

Mahoney, P.M.: Students' attitudes toward STEM: development of an instrument for high school STEM-based programs. J. Technol. Stud. **36**(1), 24–34 (2010)

Phinney, J.S., Ong, A.D.: Conceptualization and measurement of ethnic identity: current status and future directions. J. Couns. Psychol. **54**(3), 271–281 (2007)

Shou, X., Hu, W.: A probe into the project-based STEM learning model of primary school students based on the view of culture-historical activities. China Educ. Technol. **2**, 33–41 (2017)

Zhan, Z., Li, K., Lin, Z., Zhong, B., Mai, Z., Li, W.: Interdisciplinary education for cultural inheritance (C-STEAM): 6C model and its typical case. Mod. Distance Educ. Res. **32**(2), 29–38+47 (2020)

Zhan, Z., Mei, H., Mai, Z., Shao, F.: Creative thinking and innovative thinking: connotation analysis, linkage and prospect. Mod. Distance Educ. Res. (2), 40–49+66 (2019)

Zhan, Z., Zhong, B., Huo, L., Huang, M.: Transdisciplinary education for cultural inheritance (C-STEAM): value orientation and classification framework. China Educ. Technol. **3**, 69–76 (2020)

Zhang, S.: The research on traditional handcrafts intangible cultural heritage protective effect evaluation model construction and protection system based on the stakeholder theory. Master's thesis, Beijing Jiao Tong University (2014)

Zhao, H.: Instructional design and its enlightenment of STEM school in North Carolina Middle School. China Educ. Technol. **2**, 47–54 (2017)

Zhou, L.: The course of traditional culture rooted in local areas: a case study of Ba Yu culture in Ba Shu middle school. People's Educ. **22**, 32–35 (2016)

The Experiential Learning Unit for Promoting Students' Understanding of Vapor Pressure and Related Concepts

Siwa Mahardthai, Supan Yodyingyong(✉), Namkang Sriwattanarothai, and Piyachat Jittam

Institute for Innovative Learning, Mahidol University, Nakhon Pathom, Thailand
Supan.yod@mahidol.edu

Abstract. Vapor pressure is a difficult concept for high school students because it requires understanding at the microscopic level. To better understand the concept, students should clearly understand some prerequisite concepts, such as evaporation, condensation, and pressure. Therefore, this research aims to develop Kolb's experiential learning activities for providing experience before extending to learn the vapor pressure concept through online resources. The learning unit consists of two main activities that are 1) observing scientific phenomena aims at helping students to gain experience of the concepts, and 2) color manipulatives designing to help students to understand the concepts at the microscopic level. The learning unit had been implemented with grade ten students. The conceptual test and interview had been used to assess student understanding. Results show that some students still have learning difficulties related to vapor pressure and related concepts. They had difficulty in explaining the concepts at the molecular explanation and confused some technical terms. The online resources have been discussed for suggesting a suitable pathway to each group of students.

Keywords: Color manipulatives · Condensation · Evaporation · Experiential learning · Pressure · Vapor pressure

1 Introduction

Vapor pressure is the pressure exerted by a vapor in a dynamic equilibrium of evaporation and condensation at a given temperature in the closed system (Price 2001). This concept is a fundamental concept for learning thermodynamics. To understand the vapor pressure concept, students should understand the prerequisite concepts: evaporation, condensation, pressure, closed system, and equilibrium. However, some students held alternative concepts relating to vapor pressure, such as 1) students confuse technical terms about the state of water; droplet, vapor, mist, and 2) students do not understand the chemistry concepts at particle level (Canpolat 2006; Canpolat et al. 2006; Chang 1999; Pinarbasi et al. 2009).

Bain et al. (2014) showed that thermodynamics, including vaporization, condensation, and vapor pressure equilibrium, are difficult to comprehend because it is abstract

L.-K. Lee et al. (Eds.): ICTE 2020, CCIS 1302, pp. 13–22, 2020.
https://doi.org/10.1007/978-981-33-4594-2_2

and integrated with other disciplines. Furthermore, those contents are typically taught by traditional pedagogies such as a lecture, observing phenomena, or using chemical equations. They then recommended the three key success factors in learning thermodynamics, which are 1) using mathematical concepts and representations, 2) explanation at the microscopic level accommodation with results from an experiment, and 3) identify and solve the alternative concepts before moving to the next concepts. Also, Becker et al. (2015) found that the most challenging part of learning the vapor pressure concept is the explanation at the molecular level.

To learn the chemistry concept at the microscopic level, the well-liked model usually used and generally recommended is Johnstone's triangle model. Johnstone (1991) presented the learning model that can help students succeed in learning chemistry. The learning model is multilevel thought displaying the collaboration of levels of perspective: scientific phenomena at the macroscopic level, molecular explanation at the microscopic level, and chemical questions and formulations at the symbolic level. Focusing on the first level of thinking that is the macroscopic level, this level represents the scientific phenomena' descriptive and functional level. Subsequently, the microscopic level illustrates a mechanism or an explanation at the molecular level to help to understand the phenomena. Finally, the symbolic level is utilized to represent and communicate concepts and ideas in the chemistry context (Talanquer 2011). For students to construct chemical knowledge based on Johnstone's model, they will first describe the scientific phenomena from experiments or daily life events. Secondly, students will explain what happens and how it happens in a chemistry context using theory or scientific principles. And thirdly, students share or discuss their understanding that they gained and constructed with other students.

Kolb's experiential learning (1984) is the learning approach that can help students construct and transform knowledge through experience. It can be used to support or integrate to teach concepts with the microscopic level. This learning approach emphasized how students learned, including cognition, environmental factors, and emotions, and provides students with opportunities to construct knowledge in different experiences. Moreover, it can evaluate their understanding of the abstract concept, such as the vapor pressure concept, through different perspectives.

As mentioned earlier, vapor pressure and the associated concepts are challenging to understand, especially at the molecular level. To support the student in understanding these concepts, they need experiences by direct observation of the scientific phenomena. They should then have to interpret the phenomena and construct explanations generated based on the theory or principle at the microscopic level before learning through suggested online resources. Therefore, this research aims to develop Kolb's experiential learning activities for providing experience relating to vapor pressure and related concepts associated with Johnstone's model.

2 Methodology

2.1 Participants

The participants of this research are grade 10 students who study science and mathematics. They were invited to the study because they already learned the essential concepts of vapor pressure such as atomic structure, chemical bonding, molecular forces, and state of matter. Thereby, sixteen students volunteered to perform in the research activities. The learning activities require 3 h to implement and be distinguished into various sections for students' learning according to activities designed in the classroom.

2.2 Research Design

This research is designed as a one-group pretest-posttest design (see in Fig. 1). An open-ended question and a semi-structured interview were used to collect data to see students" conceptual understanding of vapor pressure and related concepts.

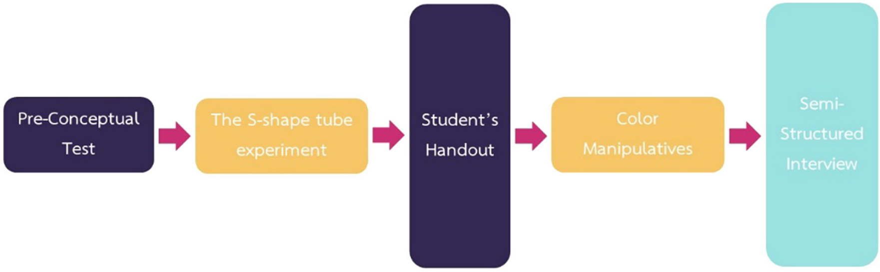


Fig. 1. The research design of this study.

2.3 Learning Activities

The learning activities consist of two parts. The first part aims to help students gain experience of the concepts by observing the S-shape tube experiment (scientific phenomena related to pressure, evaporation, and condensation) at the macroscopic level. The second part is designed to help students understand the vapor pressure and associated concepts at the microscopic level using the color manipulative to transfer their idea into the concrete explanation.

2.3.1 The S-Shape Tube Experiment

The first activity aims at helping students gain direct experiences of vapor pressure and associated concepts by observing the phenomena of evaporation, condensation, pressure, and vapor pressure at the macroscopic level via S-tube experiments. The S-shape tube experiments was developed to promote understanding of the vapor pressure concept (Brummel et al. 1959; Chen et al. 2016; DeMuro et al. 1999; Levinson 1982; Papai et al. 2019). Moreover, the experiment was designed to promote the integration of all essential and necessary concepts by linking each idea to vapor pressure. Each of the portions of setting up, the events, namely evaporation, condensation, pressure, and equilibrium between evaporation and condensation, are presented and to use the activity's handout to guide students for perceiving the vapor pressure in Fig. 2. Students

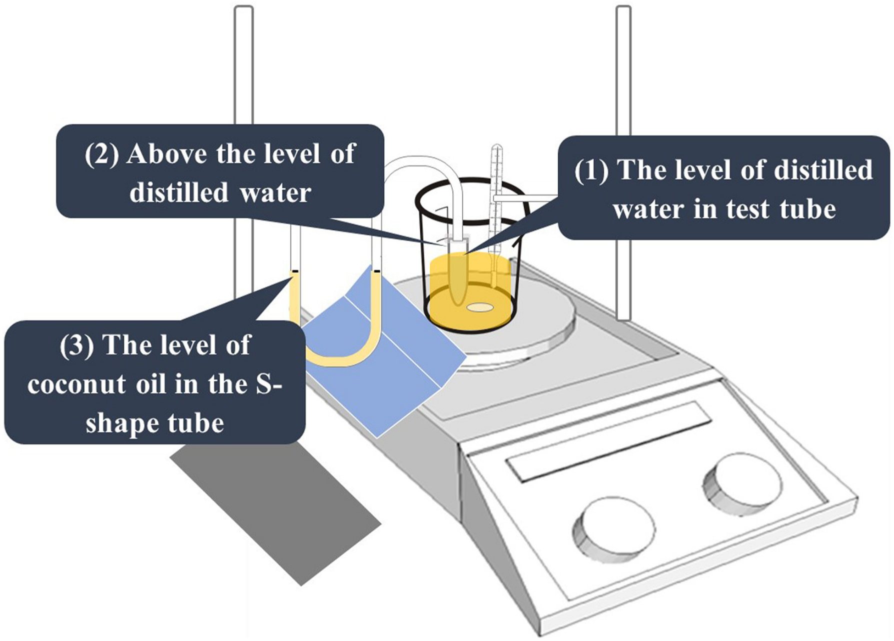


Fig. 2. The S-shape tube experiment that requires students to observe. (1) the level of distilled water and scientific phenomena in test tube, (2) the scientific phenomena that happen above the level of distilled water, and (3) the level of coconut oil in the S-shape tube.

are required to explore and construct each concept's explanations during experimenting using the students' handouts.

Figure 2 represents the observation points in the S-shape tube experiment in which the scientific phenomena happened.

(1) The level of distilled water in the test tube represents the evaporation required to detect the level of distilled water changes.
(2) Above the distilled water level: which the condensation is identified by detecting the water droplets.
(3) The coconut oil level in the S-shape tube: which the water vapor exerts to change coconut oil level as the pressure of water vapor.

2.3.2 The Color Manipulatives' Activity

The second activity aims to help students to understand the vapor pressure and related concepts at the microscopic level. The color manipulative was used as a tool to support students' learning during the explanation of each concept. Students are required to present their understanding from the first activity by generating the molecular model to explain and share ideas using the colored buttons to represent atoms or molecules.

Figure 3 presents an example of a student's model to demonstrate the process of evaporation as a molecular explanation. Students are required to utilize the colored buttons in this activity. Not only students used it to present their ideas, but teachers also used the color manipulatives to evaluate student's understanding. Additionally, teachers could adjust incorrect ideas to hold the alternative concepts to be the correct ideas after activities finished.

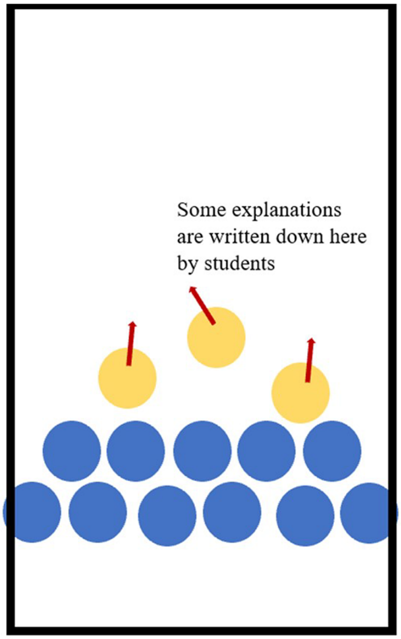


Fig. 3. The color manipulatives' activity that requires students to construct.

2.4 Data Collection and Analysis

According to this research's purpose, the assessment tools are established to determine the conceptual understanding of the vapor pressure. It is developed based on previous studies' alternative concepts (Canpolat 2006; Chang 1999; Pinarbasi et al. 2009). In addition, the content validity of the test questions was assessed by chemistry lecturers. Tools for data collection consists of four tasks as detailed:

(1) Pre-conceptual test was designed based on an open-ended question to collect information about students' prior knowledge concerning vapor pressure and related concepts.
(2) Student's handouts were the worksheets with questions related to the S-shape tube experiment procedure.
(3) Color manipulatives' handout was used to assess the in-depth perception of vapor pressure and related concepts.

(4) Interview questions was a semi-structured interview for collecting data to clarify and confirm students' responses in the handouts, and to examine their conceptual understanding of those concepts associated with color manipulatives to reveal and transform the mental model

To analyze the responses to an open-ended question in pre-conceptual test and students' handouts, each item's possible solutions with the given points are pre-determined and validated by the experts and used as criteria for scoring. Additionally, a thematic analysis is used to seek the understanding of each vapor pressure concept. As the interview results, the conversation between interviewers and the respondents was analyzed using a thematic approach to discover understanding in each concept.

3 Results and Discussion

3.1 Pre-conceptual Test and Students' Handout

The purpose of this analysis was to identify the vapor pressure concept after learning in all activities. Students' answers were grouped and interpreted using bottom-up analysis. The results present in Table 1 show the level of expressed students' understanding of the task. In addition, the description of each level of understanding is developed from Burnett (1999) and Øyehaug and Holt (2013).

Table 1. Show the categorization indicating and the number of student's responses.

Level	Description		Number of students ($n = 16$)
0	No understanding	Any explanation of vapor pressure and related concepts do not express in the task	0
1	Specific alternative concept	Answers that present the incorrect explanation in the chemistry context or show the wrong technical terms in the task	2
2	Partial understanding with specific alternative concept	Answers that present some relevant independent aspects of the concept, but some incorrect ideas and wrong technical terms are presented in the task	10
3	Partial understanding	Answers that present correctly some relevant independent aspects of the concept using the correct technical terms in the task	4
4	Sound understanding	Answers that present the collaborated relevant independent aspects and complete ideas correctly of the concept using the correct technical terms in the task	0

Table 1 displays the number of students who were assessed the understanding of vapor pressure after performing the learning activities. Overall, it is noticeable that ten students are in the level 2 presenting a partial understanding with the specific alternative concepts. For example students could explain some concepts about vapor pressure but they used incorrect technical terms or missing some subconcepts. Some students mentioned that *"Evaporation is the process that liquid phase is changed into the gas phase"*. When he was asked to provide more reasons to support how it happens, he could not express fully understand that concept. Furthermore, some students presented the alternative concept (Garnett et al. 1995), such as he mentioned the components of gas in the bubble when heated the system is oxygen gas.

However, four students have been classified into level 3 that they could answer correctly without the incorrect technical terms. Still, the ideas of vapor pressure did not show the integration as of the partial understanding. Some students represented the independent aspects of vapor pressure to answer in the task. Besides, their answers were complete and correct the explanation of vapor pressure, but it did not explain the whole idea of vapor pressure. An example of a student's response, *"Boiling is the process that liquid at room temperature gained the energy until reaching the boiling point; it changed into vapor"*.

There are two students identified at level 1 of understanding. They expressed the incorrect explanation of vapor pressure and showed more alternative concepts. One student answered that *"The vapor water loses the energy to the environment in the process of changing state of water. Then, the water droplets appear"*.

Interestingly, no students were categorized into level 4, which is the sound understanding, complete, and correct the explanation of vapor pressure.

3.2 Interview Results

The conversation between the interviewer and students was recorded to explore understanding at the microscopic level associated with color manipulatives as a tool for transferring mental perception into the concrete model. The interview results showed that ten students (level 2) could not clearly explain and construct the vapor pressure model at the molecular level. Even the interviewer added the guided question to lead them to recall the observed result; they could only show the explanation at the macroscopic level. Here is an excerpt from a student:

> *"When the distilled water heated to reach the boiling point, the evaporation occurred, and the distilled water changed state into the gas state as white smoke. The water vapor went directly to exert the coconut oil level, then the coconut oil level changes. This process is vapor pressure."*

On the other hand, the students (level 3) presented the vapor pressure explanation clearly; however, the relation of each aspect did not emerge in this group. This is because students could not construct the overall process to explain the vapor pressure concept by representing at the microscopic level. The guided questions were required to lead them for catching the molecular explanation roughly. The following quote displays this point:

> *"To change the level of coconut oil in the test tube, pressure in the experiment had to produce from the evaporation of distilled water. Besides, some molecule of*

distilled water was still the gas state that could exert to the level of coconut oil, whereas some molecule condensed into the liquid state at above the level of the distilled water in the test tube."

3.3 The Suggested Online Resources

Based on the results, the level 1–3 of expressed understanding is assessed to show the main problem of vapor pressure that lacks the molecular perception. Although students described the scientific phenomena at the macroscopic level, the explanations at the microscopic level required the explanation using theory or principle to express their answer or created model. Therefore, molecular visualization is an efficient way to provide ideas to support learning at the microscopic level (Jones et al. 2005; Kelly 2014). The online resources should illustrate the molecular behavior and the description such as Kelly (2014) utilized the molecular visualization to simulate the molecules' action for teaching in the classroom.

4 Conclusion and Implication

This research aims to develop the learning unit to promote understanding of vapor pressure using Kolb's experiential learning model for direct experience providing before learning in the online resources. This research reflects the general view that students still have difficulty conceptualizing the prerequisite concepts of vapor pressure, that numbers of students hold several alternative concepts. It could be summarized as follows:

(1) The technical terms are the main barrier of those concepts because students cannot use the words correctly, such as vapor, steam, smoke, or droplet.
(2) The scientific phenomena from learning activities are apparent; however, some students could not describe each activity's ideas.
(3) The explanation and chemical reason are rarely mentioned because almost all the students did not understand in-depth.
(4) The color manipulatives can be utilized to identify students' alternative conceptions.

As a result, the group of students who have partial understanding is suggested to explore further vapor pressure and associated concepts, particularly the idea at the microscopic level. The rationale behind this is that the color manipulatives had limited movement and were hard to represent at the microscopic level. In further study, students will learn in the online resources that can express the molecular animation and present the overview of vapor pressure. On the other hand, specific alternative concepts and partial understanding with specific alternative concepts are advised to solve the prerequisite concepts first. To elaborate their understanding of the vapor pressure concept, students should perform and observe the molecular explanation while constructing the correct ideas. The type of online resources appropriate for these groups is the resources that have the molecular animation associated with the text-based explanation.

Acknowledgements. This research was supported by Dr. Parames Laosinchai, who played roles in the construction and development of this work. He read the manuscript and provided helpful comments. Any suggestions, findings, and conclusions or recommendations expressed in this material are those of the authors. Moreover, the funding support is from the Institute for the Promotion of Teaching Science and Technology (IPST).

References

Bain, K., Moon, A., Mack, M.R., Towns, M.H.: A review of research on the teaching and learning of thermodynamics at the university level. Chem. Educ. Res. Pract. **15**(3), 320–335 (2014). https://doi.org/10.1039/c4rp00011k

Becker, N., Stanford, C., Towns, M., Cole, R.: Translating across macroscopic, submicroscopic, and symbolic levels: The role of instructor facilitation in an inquiry-oriented physical chemistry class. Chem. Educ. Res. Pract. **16**(4), 769–785 (2015). https://doi.org/10.1039/c5rp00064e

Brummel, R., Wolthuis, E., Bout, P.V.: Determination of vapor pressure: a general chemistry laboratory experiment. J. Chem. Educ. **36**(10), 494 (1959). https://doi.org/10.1021/ed036p494

Burnett, P.C.: Assessing the structure of learning outcomes from counselling using the SOLO taxonomy: an exploratory study. Br. J. Guid. Couns. **27**(4), 567–580 (1999). https://doi.org/10.1080/03069889908256291

Canpolat, N.: Turkish undergraduates' misconceptions of evaporation, evaporation rate, and vapour pressure. Int. J. Sci. Educ. **28**(15), 1757–1770 (2006). https://doi.org/10.1080/09500690600779957

Canpolat, N., Pinarbasi, T., Sözbilir, M.: Prospective teachers' misconceptions of vaporization and vapor pressure. J. Chem. Educ. **83**(8), 1237–1242 (2006). https://doi.org/10.1021/ed083p1237

Chang, J.: Teachers college students' conceptions about evaporation, condensation, and boiling. Sci. Educ. **83**(5), 511–526 (1999). https://doi.org/10.1002/(sici)1098-237x(199909)83:5%3c511:aid-sce1%3e3.3.co;2-5

Chen, W., Haslam, A.J., Macey, A., Shah, U.V., Brechtelsbauer, C.: Measuring vapor pressure with an isoteniscope: a hands-on introduction to thermodynamic concepts. J. Chem. Educ. **93**(5), 920–926 (2016). https://doi.org/10.1021/acs.jchemed.5b00990

DeMuro, J.C., Margarian, H., Mkhikian, A., No, K.H., Peterson, A.R.: An inexpensive microscale method for measuring vapor pressure, associated thermodynamic variables, and molecular weight. J. Chem. Educ. **76**(8), 1113 (1999). https://doi.org/10.1021/ed076p1113

Garnett, P.J., Garnett, P.J., Hackling, M.W.: Students' alternative conceptions in chemistry: a review of research and implications for teaching and learning. Stud. Sci. Educ. **25**(1), 69–96 (1995). https://doi.org/10.1080/03057269508560050

Johnstone, A.H.: Why is science difficult to learn? Things are seldom what they seem. J. Comput. Assist. Learn. **7**, 75–83 (1991)

Jones, L.L., Jordan, K.D., Stillings, N.A.: Molecular visualization in chemistry education: the role of multidisciplinary collaboration. Chem. Educ. Res. Pract. **6**(3), 136–149 (2005). https://doi.org/10.1039/B5RP90005K

Kelly, R.M.: Using variation theory with metacognitive monitoring to develop insights into how students learn from molecular visualizations (2014). https://doi.org/10.1021/ed500182g

Kolb, B.: Functions of the frontal cortex of the rat: a comparative review. Brain Res. **320**(1), 65–98 (1984)

Levinson, G.S.: A simple experiment for determining vapor pressure and enthalpy of vaporization of water. J. Chem. Educ. **59**(4), 337 (1982). https://doi.org/10.1021/ed059p337

Øyehaug, A.B., Holt, A.: Students' understanding of the nature of matter and chemical reactions - a longitudinal study of conceptual restructing. Chem. Educ. Res. Pract. **14**, 450–467 (2013). https://doi.org/10.1039/c3rp00027c

Papai, R., et al.: Creating and experimenting with a low-cost, rugged system to visually demonstrate the vapor pressure of liquids as a function of temperature. J. Chem. Educ. **96**(2), 335–341 (2019). https://doi.org/10.1021/acs.jchemed.8b00381

Pinarbasi, T., Sozbilir, M., Canpolat, N.: Prospective chemistry teachers' misconceptions about colligative properties: boiling point elevation and freezing point depression. Chem. Educ. Res. Pract. **10**(4), 273–280 (2009). https://doi.org/10.1039/b920832c

Price, D.M.: Vapor pressure determination by thermogravimetry. Thermochimica Acta, 367–368, 253–262 (2001). https://doi.org/10.1016/S0040-6031(00)00676-6

Talanquer, V.: Macro, submicro, and symbolic: the many faces of the chemistry "triplet". Int. J. Sci. Educ. **33**(2), 179–195 (2011). https://doi.org/10.1080/09500690903386435

A Corpus-Based Analysis of Text Complexity in NMET Reading and Its Implication on Reading Instruction

Liyan Huang, Jiaying Wang, and Xiaochen Hua(✉)

Language Cognition and Assessment Centre, School of Foreign Studies, South China Normal University, No. 55, Zhongshan Avenue West, Guangzhou, Guangdong, China
frankhxc@scnu.edu.cn

Abstract. This study builds a small-scale corpus of 54 reading texts of the National Matriculation English Test from 2016 to 2019 and uses natural language processing tools, Python and Coh-Metrix, to investigate text complexity with 35 quantitative indices from five aspects, including length, readability, vocabulary, syntax and cohesion. Results reveal the major characteristics of these texts, showing they have appropriate length, similar difficulty as the texts in high school English textbooks and a higher level of lexical diversity and stem overlap among all sentences than those in textbooks. The findings offer support for more precise and effective reading instructional practice in high school from the perspectives of the WHAT and the HOW in classroom teaching.

Keywords: Text complexity · NMET reading · Reading instruction in high school · Python · Coh-Metrix

1 Introduction

Reading is an essential competence for EFL learners. It is a challenge for many Chinese EFL learners, and the reasons are two-fold. First, reading comprehension comprises the largest portion of almost every English high-stakes test in China, e.g., the National Matriculation of English Test (hereafter, NMET). Second, reading authentic English texts is the most common approach used by Chinese EFL learners in both rural and urban areas to gain access to examples of authentic English. Reading instruction is therefore an extremely important component of English teaching in schools and so is the promotion of teaching effectiveness. In order to help students learn in the best possible way, teachers should keep many points in mind while designing and implementing a reading class, e.g., setting teaching goals, selecting appropriate materials, and employing effective teaching strategies. Currently, teachers are expected to adhere to the principles and requirements published as the National English Curriculum Standards (Ministry of Education of the People's Republic of China 2018) while planning teaching. The literature on the topic of turning the Standards into reading instructional practice is quite voluminous (e.g., Wang et al. 2019); however, few studies have analyzed the NMET and drawn out any

L.-K. Lee et al. (Eds.): ICTE 2020, CCIS 1302, pp. 23–34, 2020.
https://doi.org/10.1007/978-981-33-4594-2_3

implications for classroom teaching. Consequently, in this study we focus on the reading component of the NMET and conduct a corpus-based analysis of one of the test text's features, i.e., text complexity, and put forward a series of suggestions regarding the WHAT and the HOW issues of reading instruction in high school based on the statistical results.

2 Reading and Reading Instruction

Reading as an essential competence, both its instruction and its assessment, have long been researched. By its nature, the concept of reading is goal-directed, multidimensional and developmental (Fox and Alexander 2011). Specifically, reading is defined as a meaning-making process initiated by a reader's intention to interact with texts, which requires an integration of cognitive, motivational and socio-contextual factors, and it is also a dynamic accumulative development process across a reader's lifespan (Alexander 2012; Pearson 2009). Research on second language (L2) reading "has evolved both as an extension of first language (L1) reading and as a branch of second language acquisition" (Koda 2012, p. 304). Verhoeven (2011) pinpointed that L2 reading involves a reader's linguistic knowledge (e.g., vocabulary, grammar), the proficient use of a complex set of reading skills in both cognitive processes (e.g., working memory, schema, motivation) and social processes (e.g., technology-assisted learning, communicative learning). L2 reading also relates to the text itself in that rhetoric, genre and text complexity (e.g., lexical, syntactical, semantic features) may influence a reader's understanding (Hornof 2008). In this complex process, L2 reading involves three major operations: (a) extracting linguistic information from written language, (b) decoding the extracted information into words, sentences and paragraphs, and (c) mapping the new information onto prior knowledge (Koda 2012).

EFL learners' reading competence may gradually improve through chronic practice, with individual readers, especially struggling learners, requiring customized reading instruction to facilitate the development of their reading comprehension (Pearson 2009). Reading instruction generally includes macro reading models and micro reading strategies. Reading strategies are those deliberate, goal-directed skills such as inferencing, summarizing and comparing that attempt to support the reader's efforts to recognize words, process sentences and construct meanings out of text (Paris et al. 1984). Reading models are packages or structures of a series of reading strategies in a specific order. Popular reading models such as the top-down model, the bottom-up model, the interactive model, etc. have long been investigated to help students improve their reading competence (Shrum and Glisan 2010). Implementing such a reading pedagogy, however, does not necessarily guarantee every student success in EFL reading competence.

As reading comprehension is the consequence of an extended amount of engaged reading (Guthrie et al. 2013), selecting reading materials that will engage EFL readers plays a pivotal role in both teaching and self-learning. The criteria or guidelines for selecting such readings for EFL readers stress factors such as the reader's prior knowledge, topics of interest, cultural differences, value of information and text complexity (Guthrie et al. 2013). Especially for intermediate level readers or middle school students, there is also the factor of text complexity which is intrinsic to reading text in a foreign language (Laufer and Nation 2012). Therefore, to deepen our understanding of L2 reading and to contribute to the development of EFL reading models, it is necessary to draw on implications from the analysis of text complexity in reading materials. In light of the widely noted contribution made by the study of syntactic complexity to the development of writing instruction (e.g., Lu 2017), the current study is an attempt to propose modifications to the models of contemporary reading instruction based on our findings of text complexity.

3 Text Complexity

Broadly speaking, text complexity refers to the inherent difficulty of reading and comprehending a text combined with consideration of reader and task variables (NGA/CCSSO 2010). In a narrower sense, it is conceived as textual elements or factors that can be analyzed, studied or manipulated (Mesmer et al. 2012). In the current study, we use this term in its narrower sense.

Text complexity has been considered a critical task-related variable that may affect test-takers' reading performance, along with other variables such as genre (e.g., narratives vs. informational tests) (Gardner 2004) and topic (e.g., a more vs. less familiar topic) (Pulido 2004).

Previous studies of the effect of text complexity have examined three features of individual texts, namely words, syntax and discourse structure (Mesmer et al. 2012). Empirical studies have documented a strong relationship between vocabulary knowledge and comprehension, finding that vocabulary difficulty has influence on the understanding of text by L1 readers as well as L2 readers (Wright and Cervetti 2016). Regarding syntax, researchers have considered the ability to parse sentence syntax an essential component of reading comprehension (e.g., Graesser et al. 1996). It has been reported that the opacity and heaviness of the constituent structure of sentences make texts harder to process because readers might have difficulty in parsing the sentences (Berman 1984). Regarding discourse structure, previous investigations of text complexity have been limited to text length and cohesion. Hiebert (1999) observed that text length was one of the features that distinguished between texts of different difficulty. Several studies have reported that increasing the cohesion of a text makes it easier for readers to comprehend it (Graesser et al. 2003).

Regarding studies of text complexity of NMET texts set for reading, previous corpus-based studies have mainly examined lexical complexity. Jin et al. (2017) investigated the lexical profiles of reading texts in the NMET on a large scale, using Nation's (2006) fourteen 1K word-family lists based on the British National Corpus. They established two sets of benchmarks using the 95% and 98% text coverage targets, which provides implications for selecting and adapting reading texts in high-stakes tests. Making use of the corpus tools Wordsmith and Treetagger, Wang (2018) reported the lexical complexity of reading comprehension texts set by NMET (Jiangsu Province version) from three aspects, i.e., type-token ratio, vocabulary distribution and function words, finding that more attention should be paid to the width of vocabulary, lexical chunks and function words. Other studies have evaluated the quality of the test design of NMET's reading section by investigating text complexity on the basis of text length and readability (e.g., Peng et al. 2017). Nevertheless, the text complexity of the reading section in NMET has not yet been examined fully, especially for syntax and discourse structure. With the current study we aim to fill these research gaps.

4 The Present Study

In this study we investigate the text complexity of the reading component in the NMET. Specifically, we seek to address the two research questions: What are the characteristics of the text complexity of the NMET's reading section? WHAT does the analysis of text complexity tell us and HOW should reading instruction at high school change?

4.1 The Corpus

The current study built a small-scale corpus of 54 reading texts of the NMET, including a nationwide version and a local version (i.e., Zhejiang Province version) for the academic years 2016 to 2018.

4.2 Data Collection and Analysis

We analyzed the text complexity of reading texts in the NMET corpus with 35 measures of text length, readability, lexical complexity, syntactic complexity and cohesion, using the natural language processing tools Python and Coh-Metrix (see below). We chose these features because they have been reported to have an effect on the reading process of language learners (Alderson 2000) and on the performance of test-takers in the testing context (Bachman and Palmer 1996). These measures comprehensively reflect the text complexity of reading texts in vocabulary, syntax and discourse, providing a theoretical framework for the study of text complexity.

As mentioned above, we used two tools, Python and Coh-Metrix, because they are complementary: Python for the analysis of text length, readability and lexical complexity and Coh-Metrix for the analysis of syntactic complexity and cohesion of texts (McNamara et al. 2014). The advantage of using Python was that it integrated various functions of text normalization (e.g., cleaning texts, removing special characters) and textual analysis (e.g., counting words, computing formulas), while Coh-Metrix provided a series of reliable measures of syntactic complexity and cohesion of texts, thus having advantages at syntax and discourse levels. In addition, we included Coh-Metrix L2 readability for the reason that it was designed especially for L2 learners. The measures of text complexity and procedure were as follows (Table 1).

Text length was assessed via six measures. Before the counting of text length, texts were normalized through Python, for instance, via text tokenization, text cleaning, removing of special characters and case conversions.

Readability was assessed via three measures, including two traditional readability formulas based on sentence length and average number of syllables in words, as well as an L2 readability formula based on word frequency, syntactic similarity and referential cohesion.

Table 1. Measures of text complexity

Category	Measures
Text length	Number of words (excluding numbers or punctuation)
	Number of sentences
	Number of paragraphs
	Length of words (average number of letters in words)
	Length of sentences (average number of words in sentences)
	Length of paragraphs (average number of sentences in paragraphs)
Readability	Flesch Reading Ease
	Flesch-Kincaid Grade Level
	Coh-Metrix L2 Readability
Lexical complexity	Lexical density (ratio of content words to function words)
	Lexical diversity (type-token ratio)
	Vocabulary coverage of the British National Corpus (1K word family)
	Vocabulary coverage of the British National Corpus (2K word family)
	Vocabulary coverage of the British National Corpus (3K word family)
	Vocabulary coverage of the NMET syllabus glossary
Syntactic complexity	Number of words before main verb
	Number of modifiers per noun phrase
	Minimal Edit Distance based on part of speech
	Minimal Edit Distance based on all words
	Minimal Edit Distance based on lemmas
	Sentence syntax similarity between adjacent sentences
	Sentence syntax similarity of all combinations
Cohesion	Noun overlap between adjacent sentences
	Argument overlap between adjacent sentences
	Stem overlap between adjacent sentences
	Content word overlap between adjacent sentences
	Noun overlap of all sentences
	Argument overlap of all sentences
	Stem overlap of all sentences
	Content word overlap of all sentences
	LSA overlap between adjacent sentences
	LSA overlap of all sentences

(*continued*)

Table 1. (*continued*)

Category	Measures
	LSA overlap between adjacent paragraphs
	LSA Given-New
	Incidence score (occurrence per 1,000 words) for all connectives

Lexical complexity was assessed via six measures. Among them, vocabulary coverage of the British National Corpus (BNC) refers to the ratio of words belonging to the first 3K word families to words in the text and vocabulary coverage of the NMET syllabus glossary refers to the ratio of words belonging to the NMET glossary to words in the text. To be specific, vocabulary coverage of the BNC was calculated on the basis of three vocabulary levels of the fourteen 1K word-family lists based on the BNC 14K lists (Nation 2006), namely 1K, 2K and 3K word families, eliminating the effect of proper nouns such as personal names or place names. The BNC 14K list was based on the concept of word family, which was deemed suitable as a unit of measurement for the vocabulary requirements of receptive skills such as listening and reading (Nation and Beglar 2007). In order to maintain consistent statistical results, to calculate the vocabulary coverage of the NMET syllabus glossary we established a corresponding relation between the NMET syllabus glossary and the BNC 14K list on the basis of word family, eliminating the effect of proper nouns.

Syntactic complexity, that is, the degree of sophistication and variation of the structures produced, has been operationalized in many ways (Lu 2017). We adopted seven indices incorporated in Coh-Metrix. A higher value in the first four measures is associated with a higher degree of syntactic sophistication, whereas a higher degree of sentence syntax similarity is associated with a lower degree of syntactic variation (McNamara et al. 2014). Besides, a combination of semantic and syntactic dissimilarity measuring the uniformity and consistency of sentence construction in the text was based on the notion of a Minimal Edit Distance (McCarthy et al. 2009). It calculates the average minimal edit or the distance that parts of speech, words or lemmas are from one another between consecutive sentences in a text.

Cohesion features are explicit characteristics in a text that help create cohesive links between ideas and clauses (McNamara et al. 2014). We assessed the cohesion of the texts using eight referential cohesion indices, four Latent Semantic Analysis (LSA) indices and one incidence score in Coh-Metrix. Referential cohesion refers to overlap in content words between local sentences, or coreference, which is a linguistic cue that can aid readers in making connections between propositions, clauses and sentences in their understanding (Halliday and Hasan 1976). Referential gaps can occur when the words or concepts in a sentence do not overlap with other sentences in the text, affecting readers' comprehension and reading time according to their abilities (O'Reilly and McNamara 2007). LSA provides measures of semantic overlap between sentences or between paragraphs (Landauer et al. 2007). The LSA indices range from 0 to 1, with a higher value indicating greater cohesion (McNamara et al. 2014). Coh-Metrix also provides an incidence score (occurrence per 1,000 words) for all connectives, which play an important role in the cohesive links between ideas and clauses and provide clues about text organization (Cain and Nash 2011).

In addition, given the fact that no systematic or complete benchmarks were proposed for all the measures of text complexity, some of the measures used the values of a small-scale corpus of high school students' textbooks for reference. In order to establish the benchmarks of the textbook corpus, two experienced teachers were invited to select reading texts with similar difficulty as set by the NMET from high school students' English textbooks (People's Education Press version), including one text of practical writing, one of expository writing, one of narrative writing and one of argumentative writing. The text complexity data of 35 measures covering text length, readability, lexical complexity, syntactic complexity and cohesion were obtained through the descriptive statistics of SPSS (23.0) and used as benchmarks in a comparison of the NMET reading text corpus to determine the level of text complexity of the NMET, which is discussed in the next section.

5 Results and Discussion

The results of text complexity were as follows (Table 2).

For text length, according to the NMET syllabus and the text length requirements in reading tests put forward by scholars such as Alderson (2000), the reading texts of the NMET were of relatively moderate text length.

Based on the result, the text difficulty of the reading section in the NMET was near Grade 10 level of native speakers. Compared with the L2 Readability of the textbook corpus ($M = 16.70$, $SD = 8.25$), the reading texts of the NMET were more difficult than the textbooks for high school students.

As regards lexical complexity, the Mann-Whitney U test showed that the lexical density of NMET reading texts was not significantly different from that of the textbooks ($p = 0.41$) but the lexical diversity of the NMET reading texts was significantly higher than that of the textbooks ($p = 0.00$). The vocabulary coverage of BNC 3K word families was higher than 95%, indicating that a large number of words used in the reading texts of the NMET were high-frequency words and the texts would not present a heavy burden for test takers' reading (Laufer 1989). Furthermore, the vocabulary coverage of the NMET syllabus glossary exceeded 96%, showing that the NMET reading texts were based on the NMET syllabus glossary.

Among the results of syntactic complexity, a Mann-Whitney U test showed that only one measure, the Minimal Edit Distance based on all words, of the NMET reading texts had a significantly higher value than that of the textbook corpus ($p = 0.03$); other measures showed no significant difference. It indicated that the syntactic complexity of the NMET reading texts was at a similar level as high school textbooks.

The results of cohesion features showed that, with the textbook corpus assessed via a Mann-Whitney U test, the NMET reading texts had a significantly higher value only in one measure, that is, stem overlap of all sentences ($p = 0.04$), while other measures revealed no significant difference. This showed that the levels of cohesion features of the NMET reading texts were close to those of the textbooks of high school students, implying that the cohesion features of the NMET reading texts were suitable for testing high school students.

Table 2. Results of text complexity

Measures	*N*	*M*	*SD*
Number of words	54	282.39	10.00
Number of sentences	54	16.31	2.08
Number of paragraphs	54	4.94	0.83
Length of words	54	4.62	0.40
Length of sentences	54	18.59	2.40
Length of paragraphs	54	3.52	0.45
Flesch Reading Ease	54	58.65	5.76
Flesch-Kincaid Grade Level	54	9.70	1.27
Coh-Metrix L2 Readability	54	14.51	3.56
Lexical density	54	0.83	0.34
Lexical diversity	54	0.61	0.03
Vocabulary coverage of the BNC (1K word family)	54	0.83	0.03
Vocabulary coverage of the BNC (2K word family)	54	0.93	0.01
Vocabulary coverage of the BNC (3K word family)	54	0.96	0.01
Vocabulary coverage of the NMET syllabus glossary	54	0.96	0.01
Number of words before main verb	54	3.96	0.91
Number of modifiers per noun phrase	54	0.89	0.10
Minimal Edit Distance based on part of speech	54	0.68	0.03
Minimal Edit Distance based on all words	54	0.91	0.02
Minimal Edit Distance based on lemmas	54	0.89	0.02
Sentence syntax similarity between adjacent sentences	54	0.09	0.02
Sentence syntax similarity of all combinations	54	0.09	0.01
Noun overlap between adjacent sentences	54	0.32	0.10
Argument overlap between adjacent sentences	54	0.47	0.09
Stem overlap between adjacent sentences	54	0.42	0.10
Content word overlap between adjacent sentences	54	0.08	0.02
Noun overlap of all sentences	54	0.28	0.09
Argument overlap of all sentences	54	0.40	0.07
Stem overlap of all sentences	54	0.36	0.11
Content word overlap of all sentences	54	0.06	0.01
LSA overlap between adjacent sentences	54	0.18	0.04
LSA overlap of all sentences	54	0.17	0.05

(*continued*)

Table 2. (*continued*)

Measures	*N*	*M*	*SD*
LSA overlap between adjacent paragraphs	54	0.33	0.06
LSA Given-New	54	0.29	0.03
Incidence score for all connectives	54	82.56	8.67

6 Suggestions for Reading Instruction at High School

In this study, we took the NMET reading section as an example of the reading task and analyzed it for 35 features such as text length, readability, lexical complexity, syntactic complexity and cohesion with the natural language processing tools Python and Coh-Metrix. Our results indicate that the reading section of the NMET had proper text length, meeting the requirements of the NMET syllabus and general reading tests; however, it was of somewhat greater difficulty than high school textbooks. In terms of lexical complexity, the vocabulary in the reading section of the NMET was mostly high-frequency words and the vocabulary of the NMET syllabus glossary was also mostly of high-frequency words. In terms of lexical diversity, Minimal Edit Distance based on all words of the NMET reading texts and one measure of cohesion (i.e., stem overlap of all sentences) showed higher standards than usual in the NMET reading section. Lexical density, syntactic complexity and other cohesion measures showed no significant difference with the teaching materials of high school students in China.

Based on the summarized characteristics of the text complexity of the NMET's reading section, we, thus, believed that the answer of our first question can cast light on our further inquires. That is, what and how the analysis of text complexity can enlighten the reform of reading instruction in Chinese high schools? In the followed sections, we proposed specific implications.

6.1 The WHAT

Local education departments generally suggest that schools in the same district use the same textbook. However, due to a mismatch between existing textbooks and contextual factors such as teachers' belifes of teaching, students' competence level and individual differences, it is not unusual for teachers wanting to develop self-edited materials as supplementary teaching resources when preparing lessons. Moreover, the findings of the present study indicate that there are some differences between NMET reading texts and the reading texts in the students' textbook in terms of text complexity. Therefore, a textbook cannot be the only source for daily teaching and NMET preparation. To improve students' English competence and their performance in the NMET, teachers are advised to take account of the NMET reading texts while designing classroom instruction.

Given that developing criteria for evaluating, selecting and adapting materials is fundamental, the results of the analysis of text complexity in the study served the aforementioned issue in two aspects. First of all, it is an indispensable part of designing a reading class to set teaching goals, and to select and adapt the reading texts. Teachers might refer to our framework of text complexity, which was used to analyze 35 features of complexity, such as text length, readability, lexical complexity, syntactic complexity and cohesion, to help evaluate the quality of the reading texts and then select and edit them. Traditionally, reading instruction in high school aims at explaining vocabulary, gist, details and syntax while cohesion is ignored. To reduce the students' processing burden, it is advised that teachers should include lexical complexity, syntactic complexity and cohesion in their teaching. With regard to the selection and adaption of teaching materials, teachers might refer to the statistical results of this study and develop a set of criteria. For example, the number of words in each text could be around 280, the Flesch Reading Ease score could be approximately 60, the Coh-Metrix L2 Readability score could be around 15, the vocabulary coverage of BNC 3K word families and the NMET syllabus glossary could reach 96%, lexical density could be around 1 and the incidence of all connectives per 1,000 words could be about 85.

Secondly, the study offers possible methods for the development of reading texts or banking items for formative assessment during teaching. Every natural language processing tool has its advantages. As mentioned earlier, the use of Python reduces statistical differences due to its powerful functions while Coh-Metrix shines at syntactic and discourse levels. Changes in text complexity to suit teacher and students could integrate the relative advantages of Python and Coh-Metrix, and the teacher could conduct an in-depth analysis of selected reading texts by constructing 35 measures covering five aspects: text length, readability, lexical complexity, syntactic complexity and cohesion. Teachers could build an item banking system based on these measures, making the text complexity of reading materials for teaching and testing more sound by being scientifically based.

6.2 The HOW

The washback effect of the NMET should be taken into consideration in the design of teaching. Examinations, especially high-stakes tests like the NMET, have a great impact on the teaching in Chinese high schools. Through the analysis of the text complexity of the NMET reading section, we have primarily become aware of the fact that textual features could make a difference to high school teaching goals, teaching content and teaching methods. Thus, during the teaching process, more attention should be paid to the textual features of reading texts, for instance, the number of words, sentences and paragraphs.

Secondly, the teaching of long sentences and paragraphs should be preceded by a logical analysis of their meanings. As regards vocabulary, teachers ought to put more stress on lexical diversity and strengthen students' ability to comprehend word meanings by analyzing what it means to use a variety of words in different contexts, something that has been neglected in high school teaching.

Thirdly, since this study has been demonstrated that the ability to analyze syntax and cohesion is also required in the NMET, teachers cannot ignore the need to provide an explanation for and the learning of words modifying the main verb, syntactic patterns, overlaps among words and so on.

To sum up, while the analysis of text complexity provides an objective assessment of the difficulty of reading texts, the overall difficulty of the NMET test is also affected by factors such as test takers' proficiency, familiarity with the topic, cognition and thinking ability. In any case, an objective statistical analysis of text complexity is an important step to assist teachers in their selection of teaching materials and teaching decision. Future research could explore how to use technology to analyze the reading source text and develop more specific teaching strategies for devising or selecting reading texts at different levels of complexity.

Acknowledgment. This work was supported by grants from the National Social Science Fund of China (15BYY080) to Huang, by the Chinese MOE Research Grant for Humanities and Social Sciences (20YJC880032) to the corresponding author Hua, and the Innovation Project of Graduate School of South China Normal University to Wang.

References

Alderson, J.C.: Assessing Reading. Cambridge University Press, London (2000)

Alexander, P.: Reading into the future: competence for the 21st century. Educ. Psychol. **47**(4), 259–280 (2012)

Bachman, L.F., Palmer, A.S.: Language Testing in Practice. Oxford University Press, Oxford (1996)

Berman, R.A.: Syntactic components of the foreign language reading process. In: Alderson, J.C., Urquhart, A.H. (eds.) Reading in a Foreign Language, pp. 139–156. Longman, London (1984)

Cain, K., Nash, H.M.: The influence of connectives on young readers' processing and comprehension of text. J. Educ. Psychol. **103**(2), 429–441 (2011)

Fox, E., Alexander, P.A.: Learning to read. In: Alexander, P.A., Mayer, R. (eds.) Handbook of Research on Learning and Instruction, pp. 7–31. Routledge, New York (2011)

Gardner, D.: Vocabulary input through extensive reading: a comparison of words found in children's narrative and informational reading materials. Appl. Linguist. **25**(1), 1–37 (2004)

Graesser, A.C., McNamara, D.S., Louwerse, M.M.: What do readers need to learn in order to process coherence relations in narrative and expository text? In: Sweet, A.P., Snow, C.E. (eds.) Rethinking Reading Comprehension, pp. 82–98. Guilford, New York (2003)

Graesser, A.C., Swamer, S.S., Baggett, W.B., Sell, M.A.: New models of deep comprehension. In: Britton, B.K., Graesser, A.C. (eds.) Models of Understanding Text, pp. 1–32. Erlbaum, Mahwah (1996)

Guthrie, J.T., Klauda, S.L., Ho, A.N.: Modeling the relationships among reading instruction, motivation, engagement, and achievement for adolescents. Read. Res. Q. **48**, 9–26 (2013)

Halliday, M.A.K., Hasan, R.: Cohesion in English. Longman, London (1976)

Hiebert, E.H.: Text matters in learning to read. Read. Teach. **52**(6), 552–566 (1999)

Hornof, M.: Reading tests as a genre study. Read. Teach. **62**(1), 69–73 (2008)

Jin, T., Guo, K., Mak, B., Wu, Q.: Lexical profiles of reading texts in high-stakes tests: where are the benchmarks? Int. J. Comput. Assist. Lang. Learn. Teach. **7**(1), 34–49 (2017)

Koda, K.: Development of second language reading skills: cross-linguistic perspective. In: Gass, S., Mackey, A. (eds.) The Routledge Handbook of Second Language Acquisition, pp. 303–318. Routledge, London and New York (2012)

Landauer, T., McNamara, D.S., Dennis, S., Kintsch, W.: Handbook of Latent Semantic Analysis. Erlbaum, Mahwah (2007)

Laufer, B., Nation, I.S.P.: Vocabulary. In: Gass, S., Mackey, A. (eds.) The Routledge Handbook of Second Language Acquisition, pp. 163–175. Routledge, London and New York (2012)

Lu, X.: Automated measurement of syntactic complexity in corpus-based L2 writing research and implications for writing assessment. Lang. Test. **34**(4), 493–511 (2017)
McCarthy, P.M., Guess, R.H., McNamara, D.S.: The components of paraphrase evaluation. Behav. Res. Methods **41**(3), 682–690 (2009)
McNamara, D.S., Graesser, A.C., McCarthy, P.M., Cai, Z.: Automated Evaluation of Text and Discourse with Coh-Metrix. Cambridge University Press, New York (2014)
Mesmer, H.A., Cunningham, J.W., Hiebert, E.H.: Toward a theoretical model of text complexity for the early grades: learning from the past, anticipating the future. Read. Res. Q. **47**(3), 235–258 (2012)
Ministry of Education of the People's Republic of China: National English Curriculum Standards for General High School, 2017 edn. People's Education Press, Beijing (2018)
Nation, I.S.P.: How large a vocabulary is needed for reading and listening? Can. Mod. Lang. Rev. **63**(1), 59–82 (2006)
Nation, I.S.P., Beglar, D.: A vocabulary size test. Lang. Teach. **31**(7), 9–13 (2007)
National Governors Association Center for Best Practices & Council of Chief State School Officers (NGA/CCSSO): The Common Core State Standards for English Language Arts & Literacy in History/Social Studies, Science, and Technical Subjects, Appendix A: Research Supporting Key Elements of the Standards, Washington, DC: Author (2010)
O'Reilly, T., McNamara, D.S.: Reversing the reverse cohesion effect: good texts can be better for strategic, high-knowledge readers. Discourse Process. **43**, 121–151 (2007)
Paris, S.G., Cross, D.R., Lipson, M.Y.: Informed strategies for learning: a program to improve children's reading awareness and comprehension. J. Educ. Psychol. **76**(6), 1239–1252 (1984)
Pearson, P.D.: The Roots of Reading Comprehension Instruction. In: Israel, S., Duffy, G. (eds.) Handbook of Research on Reading Comprehension, pp. 3–30. Taylor & Francis, New York (2009)
Peng, Y., Gu, X., Huang, J.: The research on the test-designed quality of the reading comprehension among test papers of NMET national version and provincial autonomous ones in 2016. Educ. Meas. Eval. **3**, 57–64 (2017)
Pulido, D.: The relationship between text comprehension and second language incidental vocabulary acquisition: a matter of topic familiarity? Lang. Learn. **54**(3), 469–523 (2004)
Shrum, J., Glisan, E.: Teacher's Handbook: Contextualized Language Instruction, 4th edn. Heinle Cengage Learning, Canada (2010)
Verhoeven, L.: Second Language Reading Acquisition. In: Kamil, M., Pearson, P., Moje, E., Afflerbach, P. (eds.) Handbook of Reading Research, vol. IV, pp. 661–683. Routeldge, New York (2011)
Wang, R.: Study on the vocabulary of reading comprehension texts in Jiangsu National Matriculation English Test from 2008 to 2017. Educ. Meas. Eval. **1**, 19–25 (2018)
Wang, Q., Qian, X., Zhou, M.: Text analysis in English language teaching—importance and methods. Foreign Lang. Educ. China **2**(2), 40–47 (2019)
Wright, T.S., Cervetti, G.N.: A systematic review of the research on vocabulary instruction that impacts text comprehension. Read. Res. Q. **52**(2), 203–226 (2016)

A Mobile Game-Based Learning Approach for Motivating Preschoolers and Primary Students in Learning Mathematics

Chok-pang Kwok(✉), Sin-chun Ng, Ching-hei Ho, So-sum Ip, and Chi-wang Daniel Kui

School of Science and Technology, The Open University of Hong Kong, Good Shepherd Street, Ho Man Tin, Kowloon, Hong Kong SAR, China
cpkwok@study.ouhk.edu.hk, scng@ouhk.edu.hk

Abstract. Learning in an interesting and funny way is a successful key for little kids to absorb new concepts. In Hong Kong, kids are studying core subjects through "assessment" as they have to prepare for their standardized tests and examinations. Most of them are studying under high pressure to obtain excellence in the assessment. If they did not get good results, they will be frustrated and will lose their interest in learning eventually. This paper introduces the use of "game-based learning" for little kids to learn simple arithmetic calculations in a funny way. By using artificial intelligence (AI) techniques, the application can provide training exercises casually to motivate kids to learn simple arithmetic during the gameplay. Kids are focusing on how to defeat the monsters to satisfy the game goal that they will work hard to solve the simple mathematical questions. Also, the application can enhance the parent-child relationship when the kids achieve the goal in each stage with their parents together.

Keywords: Online learning · Quality assurance · Student engagement · Student satisfaction

1 Introduction

The Hong Kong Government provides 12 years of free primary and secondary education to children studying in public schools (Education Bureau 2020). In addition, a "Free Quality Kindergarten Education" (Education Bureau 2016) was released from the academic year[1] 2017/18 that the government will provide financial support to all families with their children admitted to non-profit based kindergarten.

In Hong Kong, students have adopted to study through different types of "assessment" as they have to prepare for a lot of standardized tests and examinations, especially for primary students. To achieve academic excellence, kids are studying in high pressure with many extra-curricular activities. They will face with many "intrinsic" punishments,

[1] In Hong Kong, "academic year" means the period from the first day of September to the last day of August in next year.

L.-K. Lee et al. (Eds.): ICTE 2020, CCIS 1302, pp. 35–45, 2020.
https://doi.org/10.1007/978-981-33-4594-2_4

such as physical punishment from parents (Chu 2018) if their results are not good enough (Topick - Hong Kong Economic Times 2018). Hence, some children may get frustrated and lose their confidence in learning. The worst case is some children cannot afford the pressure and commit suicides (Ngan 2019). The committee on Prevention of Student Suicides by the Education Bureau (Committee on Prevention of Student Suicides 2016) stated there is over 55% of investigated cases that are related to study problem.

Moreover, the main pedagogy used in Hong Kong education is a "spoon-feeding education" instead of "student-based education". Students are forced to "absorb" all the knowledge to satisfy all the requirements released by Education Bureau (Curriculum Development Council 2014, 2017), which may not be an effective way for students to learn. Even though the educational reform has been proposed in recent years, such as acquiring information technology and activity-based education, the fundamental education system of Hong Kong remains the same. A known protocol for teacher is to complete the syllabus on time that they are hard to take care of the student affordance and the learning quality. The Legislative Council committee (2009, 2010) pointed out that the current education system remains an outdated "spoon-feeding" which focused on "memorization" and "rote". With this approach, creativity from students will be minimal.

To motivate children in learning, stakeholders should promote more innovative ways than using the traditional teaching methodology in conducting lessons as the children in this generation need more attention than before. Also, many supports and supplementary activities are proposed by applying various technologies. Lee et al. (2019) proposed a mobile game with an attractive interface in learning English vocabularies by building models with virtual Lego blocks in using AR technology. The application can let users learn English words effectively in a funny way by building blocks virtually.

In this paper, a simple game application is proposed to motivate preschoolers and junior primary education students to learn simple mathematics by completing simple tasks in an interesting game with attractive interface. In enhancing the efficiency and correctness in mental arithmetic, dynamic practicing is an effective way to realize the preferred results. Hence, the application contains an arena so the user can control the player character to collect all required resources by purifying enemies with mental arithmetic skills as the weapon. The user can then activate the attack mode to damage the stage boss or the healing mode to protect the character by answering a mathematical question involves arithmetic calculation. This process will be repeated until the stamina of the stage boss is used up. The game will be continued to the next level. The gameplay can attract the kids to complete all stages. Moreover, not just for improving the mental arithmetic skill, the application also aims to improve the parent-child relationship as parents should accompany their child to manage the game adventure.

2 Literature Review

The application will be web-based and compatible with computers and mobile devices in choosing appropriate technologies, there will be a definite advantage which satisfies the requirements in both computers and mobile devices. In the following, Sect. 2.1 will review the current technologies, and Sect. 2.2 will review similar applications available in the market currently.

2.1 Review of Technologies

Review of Programming Languages. There are many web-based technologies such as JavaScript, Hypertext Preprocessor (PHP) and Python, which can be applied to build up web applications. However, the requirements of those languages other than JavaScript are high. For example, PHP has to be operated by a server. However, JavaScript can be used to build applications in either client-server mode or standalone mode. Also, JavaScript can be applied with Hypertext Markup Language (HTML) together easily. Curran and George (Curran and George 2012) evaluated the usability of JavaScript with other techniques for building up different types of web game applications. By applying JavaScript APIs, it can enrich the functionality of the game to increase the gaming experiences.

Review of Artificial Intelligence (AI). To enhance the gameplay of the game, AI techniques can be applied to the game. Pannu (2015) explored the standard AI techniques in use, such as Neural Network, Fuzzy Logic, Evolutionary Computing, and Hybrid Artificial Intelligence. The mechanism of each technique and the related area was discussed in detail. In this paper, AI algorithm will be explored, but the AI algorithm cannot be "too clever" to defeat the player while most of the users are the kids, and they have limited calculation ability. Kids may be frustrated if the AI enemies defeat them by generating complicated questions.

2.2 Review of Existing Solutions

Zhang et al. (2004) compared the pros and cons of classroom learning and e-learning. The assessment results by using e-learning are much higher than that of classroom learning. They concluded e-learning could perform effective teaching and learning environments for both students and teachers. Moreover, the traditional classroom learning needs students to attend the class according to a specific time and a designated location. It leads to the lack of flexibility for students to learn at anytime and anywhere. Also, the instructors need to complete the teaching at a designated place and time.

Pareto (2012) proposed an application by using AR in performing arithmetic. The application mainly focuses on students with disability that provides many activities for students to cope with their difficulties. This AR application may not be suitable for mental arithmetic because our targeted users are different. Moreover, a relatively larger space is needed to run the AR application.

Roussou et al. (2006) proposed a virtual reality (VR) application in education. However, VR is not suitable for kids, especially for kids aged 3–8; it may be harmful to their eyes. The application needs to prepare ample space to perform the activity with VR googles. The VR application is not suitable for nursery and junior primary students since they may feel dizzy when they play with VR googles. Shields and Wells (N.D.) reported that using VR headset may cause headaches, eye strain, dizziness and nausea, and myopia. Sony Interactive Entertainment Inc[2] stated the health warning that their VR device is not suitable for children aged under 12.

[2] https://www.playstation.com/en-us/network/legal/health-warnings/

Baldeón et al. (2015) proposed a game application to let kids to learn mathematics through a role-playing experience. The game is exciting that the player can become a game designer to build up a series of games for other in learning fractions. The game is strategically focusing on the children who like designing games. Teacher is the supervisor to monitor the whole development. In Hong Kong, few kids are smart enough to perform the game development, especially for preschoolers. Also, parents are hard to approve their children to learn core contents in an indirect way.

3 Methodology

3.1 Application Design

Game Interface Design. Since the users are preschoolers accompanied with their parents, the game interface should be simple engaging and colorful to attract the users. Moreover, the interface design is in responsive according to the different devices in use. The application will automatically determine the operating device to provide a consistent gaming experience. Alternatively, it provides the best fit of the screen display which satisfies the requirements of the device.

Game Play Design. In each stage, there is an arena with a stage boss to be defeated by the player by collecting a number of gems (resources). In each turn, there are several monsters at different places. The player should meet all monsters in the arena to collect the resources. When the player meets a monster, an arithmetic question will be prompted, it can be an addition, a subtraction, or a mixed arithmetic question with addition and subtraction, according to the level of the game, as shown in Fig. 1. Then, the player should select the correct answer to solve the question, and the resource can be collected by purifying that monster if the player answers the question correctly. If the player gives a wrong answer, the stamina of the player will be deducted, and the monster leaves the arena immediately. When all monsters are defeated or left, that turn will be completed. In each turn, several resources can be collected only. When the player collected the required number of resources (it can be in several turns), the player can activate the attack mode to attack the stage boss by solving a relatively complicated arithmetic question, commonly in mixed arithmetic to reduce the stamina of stage boss, as shown in Fig. 2, or activate the heal mode to recover the player stamina as shown in Fig. 3. If the stamina of the player or stage boss has not been used up, next turn is prompted to collect resources then attack the stage boss until the stage boss or the player has been defeated. Then the game advances to the next stage if the stage boss is defeated or the game is over when the player is defeated. When the game is over, the overall statistics of each gameplay will be displayed, as shown in Fig. 4.

As kids may not be familiar in using the attack mode or heal mode, the game will determine the suitable mode based on the stamina of the player; if the stamina is 80% or more, only attack mode can be activated. On the other hand, only heal mode can be activated when the stamina of player is less than 20%.

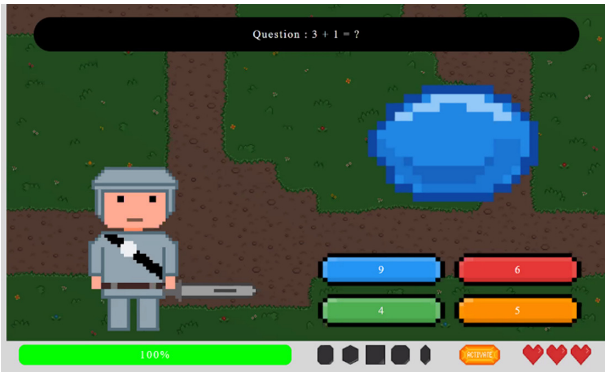


Fig. 1. An arithmetic question is prompted when the player (player) meets the monster (NPC).

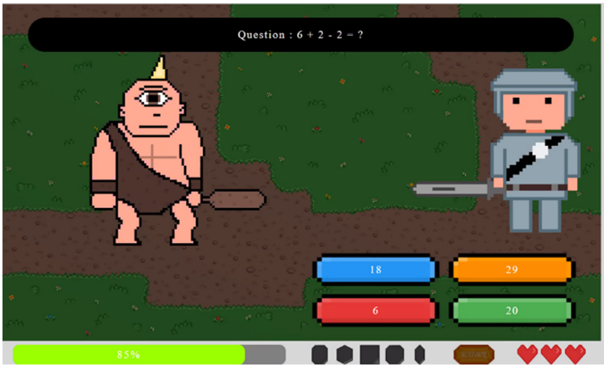


Fig. 2. The attack mode with a complicated question in defeating the stage boss.

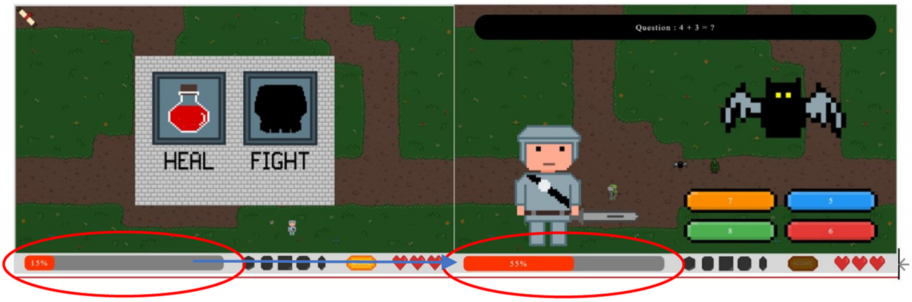


Fig. 3. The player stamina is recovered by heal mode.

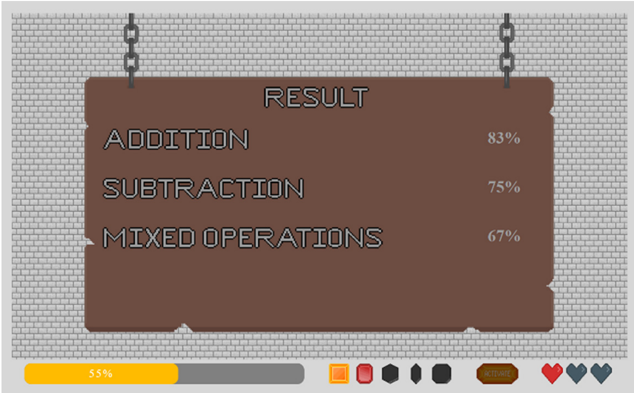


Fig. 4. The result page of the game application.

3.2 Character Movement Design

In the game, except the stage boss, all characters included player character (player, PC), and non-player characters (monsters, NPC) can move freely within the arena. There are policies to perform the best movement of the player and all monsters.

Since the game can be applied either on laptops or mobile devices, the user controls the player by keyboard may not be suitable because there is no external keyboard installed on a mobile device. Moreover, the virtual keyboard in the mobile device is not appropriate since the virtual keyboard covered around one-third of the screen, so the arena is covered then the gameplay will be affected. So, a direct position allocation mechanism is applied by using a mouse (mainly for laptops) or directly click the place of the arena (mainly for mobile devices or laptops with touch screen function) to schedule the route of character. The turning angle of the character will be calculated by using formula (1) when the player is heading to the destination. The (x_{p}, y_p) is the current coordinate of the player, and (x_{c}, y_c) is the coordinate that the player clicked on the arena. By applying the arc tangent to find out the theta (θ) from the division result of absolute axis differences which is the angle of the player should be turned like a track. The θ is calculated until the player reaches the clicked coordinate.

$$\theta = tan^{-1}\left(\frac{|y_p - y_c|}{|x_p - x_c|}\right) \tag{1}$$

The game also involved the movement of all objects except for the stage boss, finite state machine (FSM) as shown in Fig. 5, will be applied in controlling the specific movements of all monsters to maintain the best game experience. Saini et al. (2011) stated FSM is used in a fighting game to what actions will be made according to the conditions of the player. Then, performing the actions to the AI player, which simulate the gameplay that like a human player in control. In the design, there are three states which are patrol, chase and escape to control the movements of monsters. Monsters are walking around in the arena in patrol state. They find out the best route then chase the player in chase state. Then monster will leave in escape state. The escape is operated when they meet only.

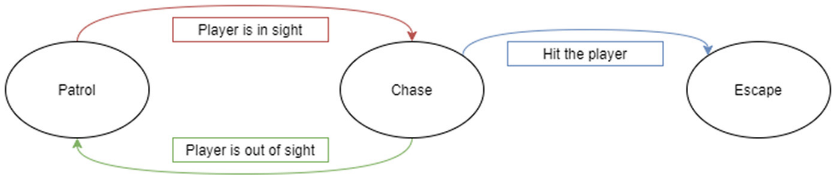


Fig. 5. States for monsters in finite state machine (FSM).

More in depth, according to the skeleton of the FSM is developed, the movement logic of each monster has been shown in Fig. 6. Monsters are generated then detect the distance to the player. If the distance is zero pixel (px) or less, which means they reached already, an arithmetic question is generated to the player. While the distance is in between 0 px and 210 px, the monster will chase the player according to the stamina of the player. If the stamina of player is more than 40%, the monster will move faster. And the monster will move slower in otherwise. While the monster is far away from the player, which is more than 210 px in the distance, the monster will generate a moving direction randomly and start to move. And the monster will get into another way if reached the arena boundary.

Moreover, the formula in (1) is also applicable to find out the turning angle of monsters. The formula in (2) is applied to find out the distance d of player and monster by the coordinates of the player with (x_p, y_p) and monster with (x_c, y_c).

$$d = \sqrt{\left(x_p - x_c\right)^2 + \left(y_p - y_c\right)^2} \tag{2}$$

3.3 Arithmetic Questions Design

The arithmetic questions design is the core part of the game. Since the target of the game is for nursery and junior primary students, and their calculation skills may not be substantial. The questions will be generated from the easiest to the hardest. That means questions will be generated an addition and subtraction with two terms in the first few stages. Then, according to the results, the complexity of the game will be increased or remain unchanged.

Relatively more straightforward questions will be generated in the beginner levels, like two terms with single digit addition (e.g. 4 + 3) and subtraction (e.g. 9-2). While the result of the scope is satisfied that the types of questions will be enhanced, like three terms arithmetic calculations (e.g. 1 + 2 + 3, 4 + 5–7). All answered questions would be recorded and analyzed. If any scope becomes weak, those types of questions are generated frequently as a remedial. However, the user may not know the type of questions that are duplicated as they are focusing on how to defeat the stage boss. For junior primary students, four terms with double digits in mixed arithmetic calculation

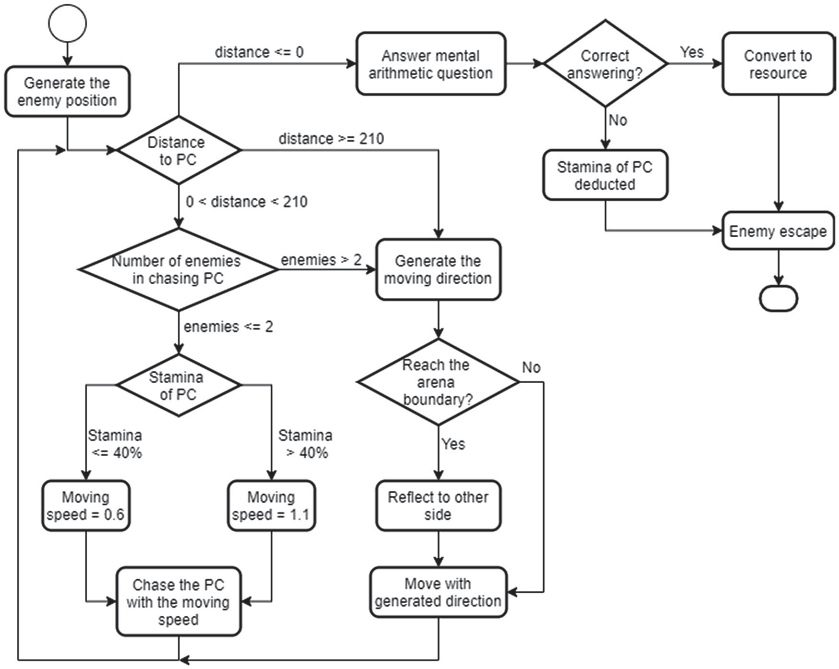


Fig. 6. The logic of monster movements.

will be the highest level in the game. For nursery students, the highest level is set to three terms with single digit mix calculated with addition and subtraction.

In details, Fig. 7 shows the logic in generating questions at a level. In each level, the maximum of possible questions will be concluded. According to the percentage of the questions generated, that means the player answered a quantity of questions, the model will start to search questions that answered incorrectly then mixed up with new questions then generate to the player by a "random factor". The random factor is a policy with a weighting value to consider generating a new question or regenerating an incorrect question from the result. The factor will tend to select incorrectly answered questions when the most of new questions are generated.

When 70% or more of new questions generated, the model will release questions according to the player's performance. If the correctness (performance) is 10% or lower, the model treated the player is weak in that scope. Remedial is needed then questions answered incorrectly that will be regenerated. When the performance is in between 10% and 80%, a new question or an incorrect question will be generated according to the random factor. Also, duplicated question cannot be selected continuously. If the performance is 80% or higher, a new question will be generated.

To make sure the stage boss can be defeated, the attack power of the player is related to the valid questions in each stage. For example, if 20 valid questions can be generated, the attack power of the player will be 5%. More valid questions can be generated in higher levels so that the attack power will be weakened. To avoid the player attends too many questions, the model will limit similar format of questions (e.g. $1 + 1$ and

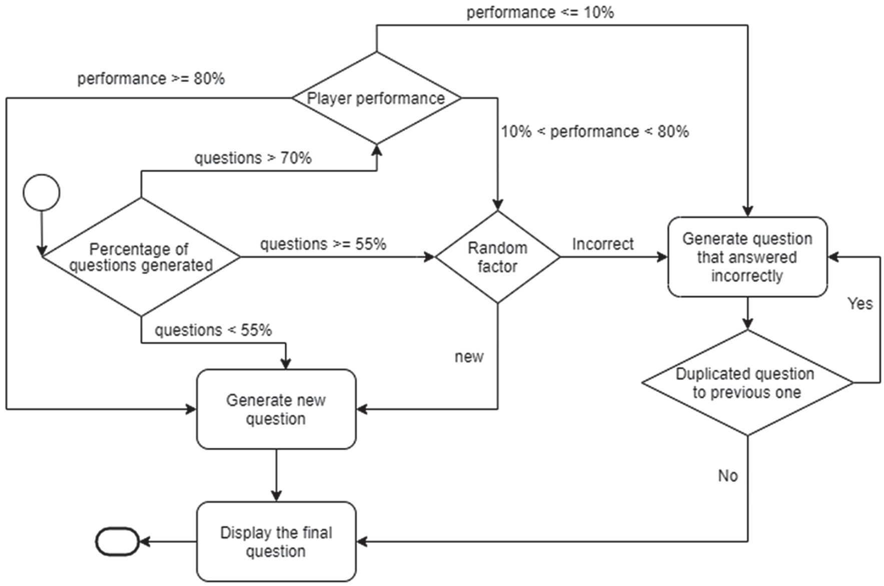


Fig. 7. The logic of question generation.

1 + 2) can be generated. Hence, the performance of the player in a stage can be more significantly performed.

4 Evaluation and Discussion

The application had been evaluated by thirteen preschoolers and junior primary children, accompanied by their parents. The evaluation was instantly followed by two post-experiment surveys, targeted for the children and their parents, respectively.

The game mode attracts all the children. They focus on how to collect the resources by applying their arithmetic ability, and how they can defeat the stage boss. Simple graphics can effectively attract their eyes to focus on the required targets.

However, the result in arithmetic improvement cannot be figured out at the moment because the evaluation period is too short. Pre- and post-tests with two sets of similar questions have been given to children to be completed before and after playing the game. Although the improvement has not been seen significantly, the reactions in doing calculation has been enhanced after completing the game.

From the feedback of the parents, the game could improve the arithmetic ability of children after playing for a period. Since the questions are at beginner level, some parents suggested the complexity can be increased and advanced arithmetic calculations can be added, for example, two digits arithmetic calculation, multiplication and division questions can be added. Furthermore, one of the parents suggested that the game can further be extended to other subjects, such as the Chinese language with word phrases matching and vocabulary spelling in the English language.

Parents feel comfortable about the game mode, they commented that it is well designed for kids as the game does not contain any violence or blood scenes. And it brings out a positive message that the evil will eventually be defeated by justice. The gameplay can effectively encourage the user to complete a task individually or work in a team with their parents. It has been frequently observed that when the children came up with a challenging question, they seek for the assistance from their parents.

The game is colorful to motivate the kids to play. Pixel-liked characters are set in high contrast that kids can see the targets clearly. Moreover, the game performed questions strategically, which focus on the weakest area of the users according to their previous performance in an individual run. This avoids the boredom brought by regular "drill and kill" methods but it can still consolidate the knowledge of the game player on weaker aspects.

In general, the application can satisfy the objectives to improve the mental arithmetic ability. Kids are motivated to communicate with their parents in learning through game playing. Hence, the parent-child relationship can be enhanced accordingly.

5 Conclusion

Learning environment for kids should be relaxing and enjoyable especially for the next generation kids. In this paper, a mobile game-based application has been developed to motivate the kids to learn mental arithmetic. The application promotes the use of game playing during learning. It can eventually enhance the users' mental arithmetic skills, problem-solving skills, and also the parent-child relationship by playing a funny and straightforward monster defeating game. The AI algorithm is set to release the mental arithmetic questions strategically to focus on the weakness of the user. As the related skill has been trained appropriately, users can experience this immersive way to learn mental arithmetic with fun effectively during game playing.

References

Baldeón, J., Puig, A., Rodríguez, I., Lopez, M., Grau, S., Escayola, M.: Gamification of elementary math learning: a game designer role-playing experience with kids. gEducation (GWC15) (2015)

Chu, E.: Ti Fa Zi Nu Shi Zui Wu Xiao De Guan Jiao Fang Fa. [Physical punishment of children is the most ineffective method of discipline]. Topick – Hong Kong Econimic Times (2018). https://topick.hket.com/article/1999446. Accessed 20 Sept 2020

Curran, K., George, C.: The future of web and mobile game development. Int. J. Cloud Comput. Serv. Sci. **1**(1), 25 (2012)

Curriculum Development Council: Basic Education Curriculum Guide. Cd.edb.gov.hk (2014). https://cd.edb.gov.hk/becg/english/index-2.html. Accessed 1 June 2020

Curriculum Development Council: Kindergarten Education Curriculum Guide. Education Bureau, Hong Kong (2017). https://www.edb.gov.hk/attachment/en/curriculum-development/major-level-of-edu/preprimary/ENG_KGECG_2017.pdf. Accessed 1 June 2020

Education Bureau: Free Quality Kindergarten Education. Education Bureau, Hong Kong (2016)

Education Bureau: Overview. Education Bureau, Hong Kong (2020). https://www.edb.gov.hk/en/student-parents/ncs-students/overview/index.html. Accessed 20 Sept 2020

Committee on Prevention of Student Suicides: Final Report November 2016. Government Printer, Hong Kong (2016)

Lee, L.-K., Chau, C.-P., Tsoi, K.-N., Yang, N.L., Wu, N.-I.: A mobile game for learning English vocabulary with augmented reality block builder. In: Cheung, S.K.S., Jiao, J., Lee, L.-K., Zhang, X., Li, K.C., Zhan, Z. (eds.) ICTE 2019. CCIS, vol. 1048, pp. 116–128. Springer, Singapore (2019). https://doi.org/10.1007/978-981-13-9895-7_11

Legislative Council committee: In Minutes of Legislative Council meeting 4 February 2019. Legislative Council: Chamber (2009). https://www.legco.gov.hk/yr08-09/english/counmtg/hansard/cm0204-translate-e.pdf. Accessed 20 Sept 2020

Legislative Council committee: In Minutes of Legislative Council meeting 18 March 2010. Legislative Council: Chamber (2010). https://www.legco.gov.hk/yr09-10/english/counmtg/hansard/cm0318-translate-e.pdf. Accessed 20 Sept 2020

Ngan, C.: Xue Ye Ya Li Chen Zhong Yi Nian Yu 10 Tong Zi Sha [More than 10 children committed suicide a year according to heavy academic pressure]. Wen Wei Po, June 2019. http://paper.wenweipo.com/2019/06/01/HK1906010026.htm. Accessed 20 Sept 2020

Pannu, A.: Artificial intelligence and its application in different areas. Artif. Intell. **4**(10), 79–84 (2015)

Pareto, L.: Mathematical literacy for everyone using arithmetic games. In: Proceedings of the 9th International Conference on Disability, Virtual Reality and Associated Technologies, vol. 9, pp. 87–96. University of Readings, Reading, September 2012

Roussou, M., Oliver, M., Slater, M.: The virtual playground: an educational virtual reality environment for evaluating interactivity and conceptual learning. Virtual R. **10**(3–4), 227–240 (2006)

Saini, S., Chung, P.W.H., Dawson, C.W.: Mimicking human strategies in fighting games using a data driven finite state machine. In: 2011 6th IEEE Joint International Information Technology and Artificial Intelligence Conference, vol. 2, pp. 389–393. IEEE, August 2011

Shields, J., Wells: Are VR headsets bad for your health? (n.d.). https://www.sciencefocus.com/future-technology/are-vr-headsets-bad-for-your-health/. Accessed 20 Sept 2020

Topick - Hong Kong Economic Times: Fu Mu Po 6 Sui Nan Tong Du Shu Du Bu Hao Da Dao Bu Neng Zi Ba Ling Ren Xin Tong [A heartbreaking issue with a 6-year-old boy was hit according to worst academic result by parents]. Topick – Hong Kong Econimic Times (2018). https://topick.hket.com/article/2101298. Accessed 20 Sept 2020

Zhang, D., Zhao, J.L., Zhou, L., Nunamaker Jr., J.F.: Can e-learning replace classroom learning? Commun. ACM **47**(5), 75–79 (2004)

Evaluating Online Tourism Education Quality by Using an Instructional Strategy Framework

Xialei Duan[1], Chung-Sing Chan[2](✉), and IpKin Anthony Wong[3]

[1] Faculty of International Tourism and Management, City University of Macau, Avenida Padre Tomás Pereira, Taipa, Macau

[2] Department of Geography and Resource Management, The Chinese University of Hong Kong, Shatin, N.T., Hong Kong

ccs_johnson@cuhk.edu.hk

[3] School of Tourism Management, Sun Yat-Sen University, Zhuhai, China

Abstract. Online education becomes increasingly popular in tourism education. This research is based on a modified instructional strategy framework for online learning environments. The framework integrates behavioral learning theory, cognitive learning theory and social learning theory and proposed seven principles to evaluate the quality of online education. Although the framework is widely applied in an online learning environment, limited research has applied this framework in tourism education. There are three main objectives in the study: (1) to evaluate the learning experience from the student's perspective; (2) to investigate the instructional strategies from the perspective of teachers; and (3) to compare whether there is any teacher-student discrepancy in the online education experience. This study adopted a mixed-method approach of data collection, which includes semi-structured in-depth interviews with teachers and a survey on the students' opinion through a questionnaire. The results indicate that feedback is heavily influencing the online course quality from the perspective of students. Meanwhile, teachers tend to consider interaction as a critical factor due to the special requirement of the tourism industry. The findings shed lights on guiding the improvements in online tourism education.

Keywords: Instructional strategy · Learning experience · Online teaching · Tourism education · Student satisfaction

1 Introduction

Online education refers to an instructional approach which facilitates teaching and learning activities in an online environment (Boelens et al. 2017; Huang 2000; Sigala 2004). It has been developed rapidly along with Internet technologies. During the Covid-19 pandemic, human mobility was severely affected (Fang et al. 2020; Yang et al. 2020). Online education immediately becomes a significant trend for most of the educational institutes in the spring semester of 2020. It enables students to participate in educational activities without physically being there (Johnson and Aragon 2003; Sigala 2002; Yan 2020).

L.-K. Lee et al. (Eds.): ICTE 2020, CCIS 1302, pp. 46–60, 2020.
https://doi.org/10.1007/978-981-33-4594-2_5

Numerous studies have been done on online instructional strategies (Cuthrell and Lyon 2007; Gaytan and McEwen 2007; Watson et al. 2017); however, the assessment on the quality of online tourism education is a highly under-researched area. The specific nature of the tourism sector requires tourism education to combine theories with practicums (Daniel et al. 2017; Zagonari 2009). The online environment is a challenge for tourism educators to incorporate both knowledge transfer and practical skill training. Therefore, it is crucial to identify the critical success factors for quality online education in the tourism field.

Although an increasing number of tourism educators are incorporating the Internet into their instruction, limited studies explored the proper instructional strategies for tourism field (Sigala 2002). Moreover, previous studies are mostly conducted from the perspective of students, and limited research was done on comparing the perception gap between students and teachers in tourism education. It is essential to understand whether what the teachers offer is what the students need in an online environment.

The objectives of this research are threefold. First, this study aims to evaluate the learning experience of tourism students in an online environment. We would like to explore the critical success factors in tourism education and how the factors would influence on student's satisfaction on the online courses. Students' perceptions can help tourism educators to understand students' needs and expectations. The second objective is to investigate the instructional strategies from the perspective of teachers. The teachers' perspective reflects the educational concept and teaching design of tourism courses. The last objective is to assess whether there is any discrepancy between students and teachers regarding online tourism courses and to put forward recommendations for improvement.

2 Literature Review

2.1 Instructional Strategies for Online Learning Environments

Instructional strategies refer to techniques that the educators adopt to motivate students, organize information and assess learning outcomes (Learning 2015). Previous scholars in the field of instructional strategies found no significant differences on student satisfaction and learning outcomes comparing the online education and classroom-based instructions (Johnson and Aragon 2003; Johnson et al. 2000). This finding led to the conclusion that technology is less important than other instructional factors, such as course content, teaching design (Phipps and Merisotis 1999). However, with the rapid development of technology, online education becomes increasingly popular. The shift from traditional instructor-led paradigm to student participation becomes both a challenge and an opportunity for online courses (Sadeghi 2019; Yan 2020). Johnson and Aragon (2003) proposed that quality online education should be based on principles that are developed from a variety of learning theories, including behavioral learning theory, cognitive learning theory and social learning theory. They further suggested a comprehensive instructional framework for an online environment, including seven principles: highlighting individual differences; motivating the student; avoiding information overload; providing a real-life context; facilitating social interaction; involving active participation, and inspiring student reflection. The seven principles in the framework incorporate 21 items in total, providing a specific guideline for online education.

This study utilized the online instructional framework to evaluate the quality of online education. Previous scholars argued that student satisfaction directly influences instruction quality. It is generally accepted that student satisfaction, which shows how positively students perceive their learning experiences, is an essential factor of learning outcomes and teaching quality (Atay and Yildirim 2009; Liao and Hsieh 2011; Scotland 2006). Therefore, this study uses student satisfaction as an indicator of online instruction quality.

2.2 Tourism Online Education

In the higher education of tourism such as Bachelor or Master's programmes in universities, classroom lecturing is usually the most common and useful mode of knowledge transfer. To supplement the practical training and enrich the experience of the students, many tourism educators have also included the part of outdoor fieldwork and field trips (Jong 2015; Leydon and Turner 2013). However, to address the disadvantages of one-way instructor-to-learners lecturing, some programmes incorporate more exercises, involvement and interaction of real-life practical context to the students (Herbert et al. 2017). Some researchers confirmed the effectiveness of active learning by applying in-class activities which are based on the theories of constructivism or social interdependence (Arthurs and Kreager 2017; Mahon et al. 2010).

Due to the impact of the Covid-19 pandemic, many in-class face-to-face lecturing had to shift to online teaching, and most of the planned field trips were also cancelled. In the context of tourism education, these field experiences were lost, but instructors had also attempted to move at least a small part of the field trips to a combined mode of video, virtual and self-paced arrangements. Such combination of real-world and virtual experience was not uncommon in both tourism industry and destination management (Davis and Singh 2015; Webster 2016), and the e-learning perspectives (Azeiteiro et al. 2015; Schott 2017). Although online teaching or e-learning might not satisfy the expectations of tourism practitioners and students (Cini et al. 2015), the teaching mechanism must be changed and transformed to online platforms unintentionally in such difficult time. Nevertheless, opportunities beyond classroom teaching could exist from the crisis because tourism education would often involve a great variety of both knowledge transfer and practical skill training. E-learning could be viewed and implemented to incorporate various teaching-learning activities that should not be constrained by time, resources, geographical settings and locations since tourism courses could cover a wide range of topics related to urban, cultural and heritage, community-based and nature-based topics and experiences (Kelner and Sanders 2009; Ting and Cheng 2017).

With multimedia and simulation of real-world situations (Callaghan 2016; Fotiadis and Sigala 2015), online teaching can incorporate a great variety of tourism context, as well as a scenario-based approach that involves instructors and students without geographical boundaries (Arthurs and Kreager 2017; Cini et al. 2015; Herbert et al. 2017; Ting and Cheng 2017). In the circumstance of the COVID-19 pandemic, many overseas instructors and students were trapped at home. It is therefore expected that online education could break the geographical limitations with the combination of multiple teaching-learning approaches, such as virtual reality (VR) (Schott 2017; Weibel et al. 2012), augmented reality (AR) (Dunleavy et al. 2009), game-based learning and e-Learning modules to attain far-reaching advantages and benefits teaching-learning experience enhancement (Benckendorff et al. 2015; Mavridis et al. 2017).

Tourism education (e.g. tourism planning and management, sustainable tourism, eco-tourism, cultural tourism, urban tourism and others) is advocated to utilize e-learning approach and turn to online mode since these areas usually involve wider global consciousness and understanding of environmental and cultural diversity (Hales and Jennings 2017; Jong 2015; Raptis et al. 2018; Xu et al. 2017). Online education in tourism study allows students to benefit from knowledge acquisition (e.g. learning of fundamental concepts, underlying theories and applying models) (Kollmuss and Agyeman 2002; Lloyd-Strovas et al. 2018; Webster 2016), attitudinal and perceptual changes (e.g. discussions about controversial topics and issues in tourist destinations) (Fatima et al. 2016; Mobley et al. 2010), as well as usability consideration (e.g. stimulation of learning and response interest, and simulation of real-world situations) (Chiao et al. 2018; Mavridis et al. 2017). However, there is still much room for empirical evidence on how online education and e-learning platform may support and improve the effectiveness and experience in teaching-learning process in tourism education (Mavridis et al. 2017).

3 Methodologies

A mixed-method approach, including both qualitative and quantitative measurement, is adopted in this study. This approach allows a comprehensive understanding of tourism online education quality from the perspective of both students and teachers. First, exploratory factor analysis was conducted to investigate determining factors for quality tourism online courses; second, regression analysis was followed to explore the relationship between the identified factors with student satisfaction towards online courses; meanwhile, teachers' opinion was collected through in-depth interviews. The perception gaps were compared at the last stage.

3.1 Design of Questionnaire and Data Collection

Student opinions were collected by using an online survey questionnaire. Section one includes instructional strategies. The items are derived from the instructional strategy framework developed by Johnson and Aragon (2003). The framework includes seven principles with 21 items in total. Section two contains one question about the satisfaction level with the quality of online courses in the spring semester of 2020. The last section of the questionnaire is an open-end question about the comments on the current online courses and suggestions for improving online learning quality. This questionnaire adopts a 5-point Likert scale, with "1" representing "strongly disagree" and "5" representing "strongly agree". The questionnaire was pre-tested with 30 students, and adjustment was made after the pilot test to ensure students' accurate understanding of each question. We selected one bachelor and one master program in tourism field from two universities in Hong Kong and Macau as our target samples. The final data set included 152 valid responses.

3.2 Data Analysis

Exploratory factor analysis (EFA) is most commonly used in reducing the dimensionality of the original set of data into new dimensions which could be generated from the old ones (Fabrigar and Wegener 2011). In this study, EFA was conducted to the 21 items in the instructional strategy framework. Boost Regression Trees (BRT) model (Carslaw and Taylor 2009) was employed after the EFA, which is developed for classification and regression trees (Elith et al. 2008; Schapire 2003). It is a statistical approach, which utilizes the technique of boosting to combine large numbers of relatively simple tree models adaptively to optimize predictive performance. BRT is capable of addressing both linear and non-linear relationships and interactions between variables. Therefore, this study adopted BTR to investigate relationships between the resulting factors generated from EFA and the satisfaction level of online courses.

3.3 Semi-structured Interview

Teachers' perspective of online education quality was investigated through a semi-structured interview, including questions about critical elements for quality online teaching and suggestions for improving online teaching quality. Face-to-face interviews with individual interviewees were conducted at the end of the spring semester of 2020, which was from 15–19 June. Five teachers from the evaluated tourism courses participated in the interviews. Two of them were from Hong Kong, and the other three interviewees were from Macau.

4 Results and Discussion

4.1 Explorative Factor Analysis

Results of the questionnaire generated information about online education from the perspective of students. Before data analysis, there is the need to ensure the reliability of the data set by conducting reliability test (Cronbach 1951, 2016). Gliem and Gliem (2003) argued that it is necessary to test the internal consistency against Cronbach's alpha when the data is in Likert scales. In this research, a reliability test was conducted on the 21 items of instructional strategies, and Cronbach's coefficient was 0.945, which demonstrated that the internal consistency of the data was quite good.

EFA was conducted after the reliability test. The Kaiser-Meyer-Olkin (KMO) measure of sample adequacy was 0.898, and Bartlett's test for Sphericity is significant. All of the 21 items of instructional strategies have remained because the loadings of all the items were above 0.4 (Cairney 2003). Four factors were identified after conducting EFA, which were named (1) "Interaction" (6 items, $\alpha = 0.857$), (2) "Course content" (8 items, $\alpha = 0.925$), (3) "Feedback" (4 items, $\alpha = 0.938$) and (4) "Course schedule" (3 items, $\alpha = 0.926$). The four factors had Eigenvalue greater than one and accounted for over 69% of the total variance (Table 1). "Interaction" refers to collaborations with classmates. Items related to teamwork, case studies, group discussions are included in this factor. "Course content" refers to course materials and activities arranged by teachers. Course formats, online platforms are the basis of course content. Teachers also try to motivate students when arranging course contents, such as encouraging student activities, incorporating games into the contents, and adopting multimedia in the course materials. "Feedback" mainly focuses on comments from both teachers and students. The last factor "course schedule" includes items that are related to the course outline and schedule. The EFA results summarized the four critical factors of instructional strategies for an online learning environment.

4.2 Boost Regression Trees

After EFA, we employed boost regression trees to explore the relationship between instructional strategies and online course quality. During the development of BRT model, four factors generated in the EFA were included as independent variables. The satisfaction of online courses was the dependent variable. Among the 152 valid responses, 137 of them were randomly chosen as training datasets while the remaining 15 was treated as test dataset. Figure 1 depicts the comparison between observed satisfaction and the results from BRT predictions. The mean squared error between test data and the predicted value is 0.495, suggesting a good regression result. Table 2 demonstrates the importance of the four independent variables. Feedback accounts for 64.2% of the influence on the satisfaction level of online courses, indicating it is the most crucial instructional strategy. Content has around 20% of the influence, which is the second important factor. Course schedule and interaction have relatively lower influence over the satisfaction level, suggesting these two factors are of low importance in online teaching.

The result demonstrates that students value the feedback most in online courses. They hope to build a personal connection with the teachers and receive timely feedback on their performance. Meanwhile, feedback from classmates is also crucial since students care about peer evaluation results. Course content is second important, which includes course materials, activities, assessment, format and online platforms. Students reported that they hope to have all the online courses in one platform instead of multiple ones. Course schedule arrangement does not have a high level of influence over course satisfaction. Students are reluctant in interaction with classmates. The reasons could be reflected from their comments. Many students mentioned that they hope to reduce group work because online communication with groupmates is not convenient. They also pointed out that some of the online platforms do not function well. Interaction and discussion online always result in a waste of time and effort. They hope to use online platforms that they are familiar with, such as social media applications. The students also showed discontent towards teamwork because some of the groupmate did not contribute at all. This phenomenon becomes even worse in an online environment.

4.3 In-Depth Interviews

Semi-structured interviews were conducted based on the four instructional strategies (feedback, content, schedule and interaction). Five teachers from the surveyed tourism courses were invited to talk about the importance of different instructional strategies in tourism education. The results generate some common findings. Most of the teachers demonstrated a positive attitude towards interaction and course content.

Table 1. EFA factor loadings

	1 interaction	2 content	3 feedback	4 schedule
1. Online course materials are provided in multiple formats/platforms		.742		
2. I can access the materials according to my time schedule		.572		
3. Teachers encourage active and collaborative work		.575		
4. Teachers incorporate games into the online courses		.533		
5. There are guest speakers in my online courses		.721		
6. Teachers use multimedia in the course material		.583		
7. Teachers arrange several breaks during an online class				.673
8. Teachers arrange classes around the course outlines				.721
9. Teachers provide a visual representation of the course structure				.703
10. Students are formed into virtual learning teams for the online courses	.784			
11. Teachers use appropriate case studies related to the course content	.646			
12. Students have course projects which requires collaborations with organizations in tourism industry	.574			
13. I can feel the personal connection with the teachers of the online courses			.700	
14. I can feel the personal connection with the classmates of the online courses	.707			
15. I receive comments and feedback from my classmates of the online courses			.714	
16. The online courses incorporate active learning through a project-based approach			.664	
17. Students are encouraged to work in teams	.721			
18. Students have small group discussions during the online courses	.797			

(*continued*)

Table 1. (*continued*)

	1 interaction	2 content	3 feedback	4 schedule
19. Teachers provide extensive and timely feedback			.578	
20. Students have test/assignment/exercises during the online courses		.525		
21. Students are required to do online diaries or reflective journals		.602		
Eigenvalue	4.419	3.570	3.457	3.130
Explained variance (%)	21.042	17.002	16.461	14.903
Accumulated variance (%)	21.042	28.044	54.505	**69.408**

Extraction Method: Maximum Likelihood. Rotation Method: Varimax with Kaiser Normalization.[a]

a. Rotation converged in 7 iterations.

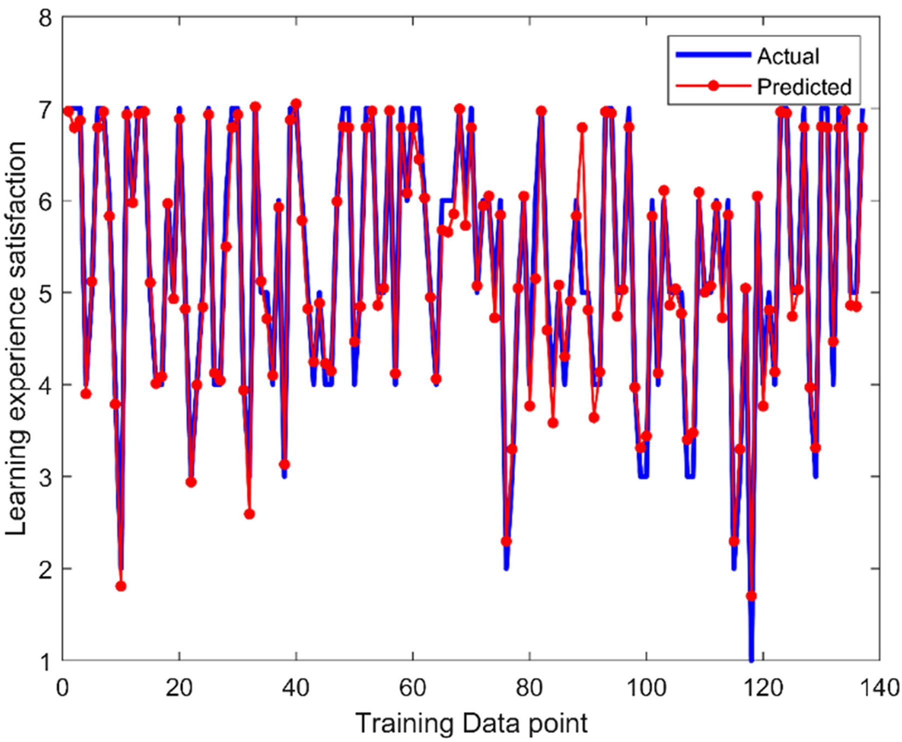
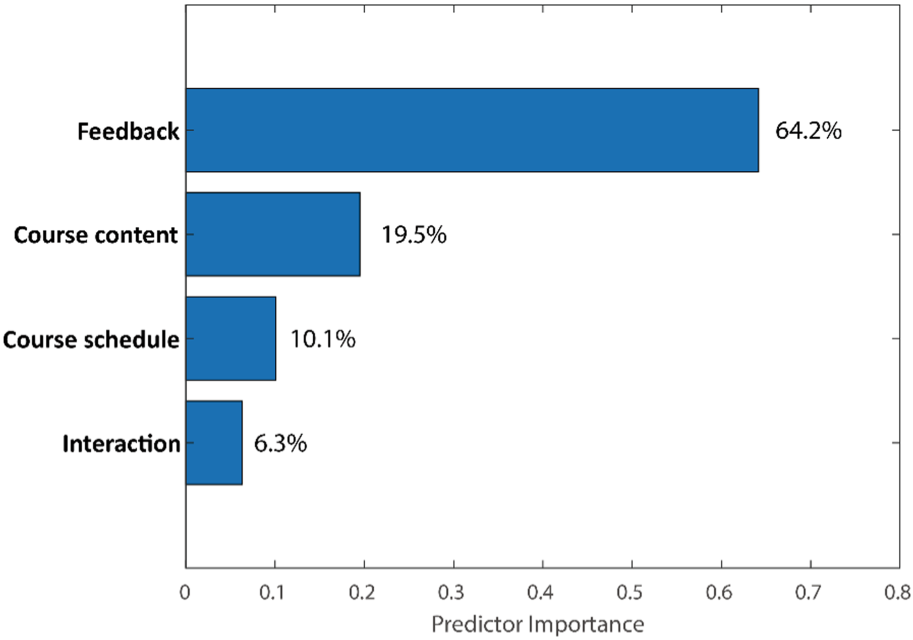


Fig. 1. Comparison between observed satisfaction and predicted value

Firstly, all of the interviewees mentioned the importance of interaction in online teaching in the tourism field since the service industry requires a workforce to have a

Table 2. Importance of instructional strategies from the perspective of students



relatively high level of people skills. They consider interaction would help students to participate in class activities actively and generate better learning outcome. One interviewee reported that it was a surprise that online interaction is even better than classroom discussion since students may feel more confident and less pressure through online instant messaging than talking in front of the whole class. Moreover, some interviewees also reported that interaction not only motivates students but also allows teachers to understand individual differences among students. In this way, teachers could adjust the course content and schedule accordingly.

Secondly, most of the interviewees consider the course content to be crucial in online teaching. It should be the main focus of the course. Many interviewees pointed out the importance of course materials. It is crucial to select proper textbooks and make adjustment according to the needs of students in different classes. Moreover, teachers consider that online learning requires a high level of self-control. Students are easily distracted by contents which are irrelevant to the course. Therefore, they all tried to make the online teaching more interesting by multiple approaches, such as inviting guest speaker, displaying fun videos, incorporating games and so on.

As for feedback, teachers consider it to be relatively less important. As reported by interviewees, "it is enough if students could actively interact and grasp the main contents during class time." The interviewees also reported that the course schedule is similar to classroom teaching. One interviewee stated that "the course schedule and outline were uploaded online at the beginning of this semester, just the same as previous semesters."

Teachers also arrange several breaks during one class, which is the same as classroom teaching.

5 Discussion and Implications

The results indicate that there is an apparent perception gap of online tourism education between students and teachers. From the perspective of students, timely feedback from both teachers and classmates were emphasized. The result echoes with Semley et al. (2016), in which they stated that providing feedback is essential in teaching and learning process. It could help enhance students' confidence when they get positive feedback from the instructor (Badami et al. 2011; Lee 1997) and stimulates intrinsic learning motivation (Badami et al. 2011). However, providing effective feedback in an online environment is a huge challenge for online education because it heavily relies on Internet technologies (Getzlaf et al. 2009). Teachers may face difficulty in providing timely feedback in an online environment. For instance, teachers may miss some comments or emails from students due to specific technical problems. Therefore, teachers should put more effort in providing effective feedback in an online environment. Furthermore, Badami et al. (2011) argued that positive feedback could help generate a higher level of intrinsic motivation than negative ones. Teachers need to offer more positive feedback to encourage students in online courses. Meanwhile, the interaction does not show significant influence over students' satisfaction with the online courses. This result is different from the previous studies, which suggests that interaction has always been an crucial factor in predicting students' satisfaction (Ali and Ahmad 2011; Bolliger 2004; Bray et al. 2008; Dennen et al. 2007; Yukselturk and Yildirim 2008). It was found that this difference is caused by the online environment, where interaction is more difficult than classroom interaction.

In-depth interview results showed that teachers generally consider interaction to be the most important factor in online tourism education since it trains the expressive ability and communication skills. Online education is shifting from teaching-centred to learner-centred, which means focusing on student participation is more important than course content (Maumbe 2014; Saulnier et al. 2008). Previous research has shown that tourism education is influenced by the dualistic nature of tourism (Paris 2011; Tribe 2008). The main goal of tourism education is to train students to work in a real business environment (Morgan 2004). Therefore, teachers should balance the vocational nature and educational contents in tourism courses in order to prepare students for the industry (Inui et al. 2006). Encouraging students' interaction in class could help students to improve their expressive ability, teamwork skills and so on. Teachers should prepare students for a future career by arranging course projects which require collaborations with organizations in tourism industry. Zehrer and Mössenlechner (2008) addressed the discrepancy between the education offered by the tourism institute and the real needs in the industry. Scholars further argued that the soft competencies such as critical thinking, communication, problem-solving skills are what the tourism industry look for in students (Daniel et al. 2017; Sisson and Adams 2013).

Although the interaction is significant in tourism courses, students showed little interest in the online courses. The main reason is that interaction does not function well when it happens in an online environment. For instance, the interaction is always interrupted by technique problems. Therefore, improving the Internet stability and Internet self-efficacy should be addressed in online courses. AlHamad et al. (2014) stated that convenient and easy navigation would increase students' interest in the online courses. Teachers needs to figure out a way to facilitate interaction in a more convenient approach, such as utilizing social media platforms. Many students reported that they feel more comfortable in using platforms that they are familiar with in their daily life.

Apart from the perception gaps, both students and teachers consider course content to be an important factor in online education. Interesting course contents and proper assessment are essential parts of successful online courses. Incorporating games and inviting guest speakers could add fun to the online class and improve student's learning motivation. However, students and teachers show a different attitude towards online platforms. Students prefer a unified online platform for all the courses since it could be easier for them to get familiar with the functions in one platform. Teachers seem to have different preferences in choosing platforms for their course. Therefore, it is suggested that universities should provide sufficient training for both students and teachers to improve Internet self-efficacy since it affects the learning process and outcome (Gangadharbatla 2008; Kuo et al. 2014; Tsai 2012).

6 Conclusion

This study has examined the online tourism education quality by modifying an instructional strategy framework. Through EFA, four factors were generated from the seven instructional strategies, which includes interaction, course content, feedback and course schedule. BRT was conducted afterwards and showed that feedback heavily influences the quality of online education. In contrast, semi-structured interviews with the teachers of the surveyed courses showed that interaction should be the most important structure in online tourism education. The result indicates that there is an apparent perception gap between students and teachers. To improve the quality of tourism online education, the educational institute should provide a stable online environment and good training for both students and teachers. The teachers need to use frequent feedback to keep students motivated. The interaction is an essential part of training students' abilities for a future career in the tourism industry; teachers need to facilitate interaction in a more convenient approach, such as utilizing social media (Callaghan 2016; Moghavvemi et al. 2017). Future research could focus on how Internet self-efficacy affect tourism online education.

References

AlHamad, A.Q., Al Qawasmi, K.I., AlHamad, A.Q.: Key Factors in determining students' satisfaction in online learning based on'Web Programming'course within Zarqa University. Int. J. Glob. Bus. **7**(1), 7 (2014)

Ali, A., Ahmad, I.: Key factors for determining student satisfaction in distance learning courses: a study of Allama Iqbal Open University. Contemp. Educ. Technol. **2**(2), 118–134 (2011)

Arthurs, L.A., Kreager, B.Z.: An integrative review of in-class activities that enable active learning in college science classroom settings. Int. J. Sci. Educ. **39**(15), 2073–2091 (2017)

Atay, L., Yildirim, H.M.: Determining factors that affect satisfaction of students in undergraduate tourism education (2009)

Azeiteiro, U.M., Bacelar-Nicolau, P., Caetano, F.J., Caeiro, S.: Education for sustainable development through e-learning in higher education: experiences from Portugal. J. Clean. Prod. **106**, 308–319 (2015)

Badami, R., VaezMousavi, M., Wulf, G., Namazizadeh, M.: Feedback after good versus poor trials affects intrinsic motivation. Res. Q. Exerc. Sport **82**(2), 360–364 (2011)

Benckendorff, P., Lohmann, G., Pratt, M., Reynolds, P., Strickland, P., Whitelaw, P.: Enhancing student learning outcomes with simulation-based pedagogies. Australas. Soc. Comput. Learn. Tert. Educ. (ASCILITE), 618–620 (2015)

Boelens, R., De Wever, B., Voet, M.: Four key challenges to the design of blended learning: a systematic literature review. Educ. Res. Rev. **22**, 1–18 (2017)

Bolliger, D.U.: Key factors for determining student satisfaction in online courses. Int. J. E-Learn. **3**(1), 61–67 (2004)

Bray, E., Aoki, K., Dlugosh, L.: Predictors of learning satisfaction in Japanese online distance learners. Int. Rev. Res. Open Distrib. Learn. **9**(3) (2008)

Callaghan, N.: Investigating the role of Minecraft in educational learning environments. Educ. Media Int. **53**(4), 244–260 (2016)

Carslaw, D.C., Taylor, P.J.: Analysis of air pollution data at a mixed source location using boosted regression trees. Atmos. Environ. **43**(22), 3563–3570 (2009)

Chiao, H.-M., Chen, Y.-L., Huang, W.-H.: Examining the usability of an online virtual tour-guiding platform for cultural tourism education. J. Hosp. Leis. Sport. Tour. Educ. **23**, 29–38 (2018)

Cini, F., Van der Merwe, P., Saayman, M.: Tourism students' knowledge and tenets towards ecotourism. J. Teach. Travel. Tour. **15**(1), 74–91 (2015)

Cuthrell, K., Lyon, A.: Instructional strategies: what do online students prefer. MERLOT J. Online Learn. Teach. **3**(4), 357–362 (2007)

Daniel, A.D., Costa, R.A., Pita, M., Costa, C.: Tourism education: what about entrepreneurial skills? J. Hosp. Tour. Manag. **30**, 65–72 (2017)

Davis, K., Singh, S.: Digital badges in afterschool learning: Documenting the perspectives and experiences of students and educators. Comput. Educ. **88**, 72–83 (2015)

Dennen, V.P., Aubteen Darabi, A., Smith, L.J.: Instructor–learner interaction in online courses: The relative perceived importance of particular instructor actions on performance and satisfaction. Distance Educ. **28**(1), 65–79 (2007)

Dunleavy, M., Dede, C., Mitchell, R.: Affordances and limitations of immersive participatory augmented reality simulations for teaching and learning. J. Sci. Educ. Technol. **18**(1), 7–22 (2009)

Elith, J., Leathwick, J.R., Hastie, T.: A working guide to boosted regression trees. J. Anim. Ecol. **77**(4), 802–813 (2008)

Fabrigar, L.R., Wegener, D.T.: Exploratory Factor Analysis. Oxford University Press, Oxford (2011)

Fang, H., Wang, L., Yang, Y.: Human mobility restrictions and the spread of the novel coronavirus (2019-nCoV) in China (0898-2937) (2020)

Fatima, J.K., Khan, H.Z., Goh, E.: Environmental knowledge and behavioural outcomes of tourism students in Australia: towards testing a range of mediation and moderated mediation effects. Environ. Educ. Res. **22**(5), 747–764 (2016)

Fotiadis, A.K., Sigala, M.: Developing a framework for designing an Events Management Training Simulation (EMTS). J. Hosp. Leis. Sport. Tour. Educ. **16**, 59–71 (2015)

Gangadharbatla, H.: Facebook me: Collective self-esteem, need to belong, and internet self-efficacy as predictors of the iGeneration's attitudes toward social networking sites. J. Interact. Advert. **8**(2), 5–15 (2008)

Gaytan, J., McEwen, B.C.: Effective online instructional and assessment strategies. Am. J. Distance Educ. **21**(3), 117–132 (2007)

Getzlaf, B., Perry, B., Toffner, G., Lamarche, K., Edwards, M.: Effective instructor feedback: perceptions of online graduate students. J. Educ. Online **6**(2), n2 (2009)
Hales, R., Jennings, G.: Transformation for sustainability: the role of complexity in tourism students' understanding of sustainable tourism. J. Hosp. Leis. Sport. Tour. Educ. **21**, 185–194 (2017)
Herbert, C., Velan, G.M., Pryor, W.M., Kumar, R.K.: A model for the use of blended learning in large group teaching sessions. BMC Med. Educ. **17**(1), 197 (2017)
Huang, H.: Instructional technologies facilitating online courses. Educ. Technol., 41–46 (2000)
Inui, Y., Wheeler, D., Lankford, S.: Rethinking tourism education: what should schools teach. J. Hosp. Leis. Sport. Tour. Educ. **5**(2), 25–35 (2006)
Johnson, S.D., Aragon, S.R.: An instructional strategy framework for online learning environments. New Dir. Adult Contin. Educ. **2003**(100), 31–43 (2003)
Johnson, S.D., Aragon, S.R., Shaik, N.: Comparative analysis of learner satisfaction and learning outcomes in online and face-to-face learning environments. J. Interact. Learn. Res. **11**(1), 29–49 (2000)
Jong, M.S.Y.: Context-aware geography field trip with EagleEye: teachers' first experience. In: Chang, M., Li, Y. (eds.) Smart Learning Environments. LNET, pp. 77–93. Springer, Heidelberg (2015). https://doi.org/10.1007/978-3-662-44447-4_5
Kelner, S., Sanders, G.: Beyond the field trip: teaching tourism through tours. Teach. Sociol. **37**(2), 136–150 (2009)
Kollmuss, A., Agyeman, J.: Mind the gap: why do people act environmentally and what are the barriers to pro-environmental behavior? Environ. Educ. Res. **8**(3), 239–260 (2002). https://doi.org/10.1080/13504620220145401
Kuo, Y.-C., Walker, A.E., Schroder, K.E., Belland, B.R.: Interaction, Internet self-efficacy, and self-regulated learning as predictors of student satisfaction in online education courses. Internet High. Educ. **20**, 35–50 (2014)
Learning, A.: Health and life skills guide to implementation (K–9): instructional strategies. Accessed 6 June 2015
Lee, A.M.: Contributions of research on student thinking in physical education. J. Teach. Phys. Educ. **16**(3), 262–277 (1997)
Leydon, J., Turner, S.: The challenges and rewards of introducing field trips into a large introductory geography class. J. Geogr. **112**(6), 248–261 (2013)
Liao, K.-H., Hsieh, M.-F.: Statistic exploring the casual relationships between service quality, brand image, customer satisfaction and customer loyalty on the leisure resort industry. Paper presented at the Service Management (IRSSM-2) in 2011 proceedings of the 2nd International Research Symposium (2011)
Lloyd-Strovas, J., Moseley, C., Arsuffi, T.: Environmental literacy of undergraduate college students: development of the environmental literacy instrument (ELI). Sch. Sci. Math. **118**(3–4), 84–92 (2018)
Mahon, J., Bryant, B., Brown, B., Kim, M.: Using second life to enhance classroom management practice in teacher education. Educ. Media Int. **47**(2), 121–134 (2010)
Maumbe, K.: Teaching and learning in recreation and tourism: a comparison of three instructional methods. J. Teach. Travel. Tour. **14**(4), 365–385 (2014)
Mavridis, A., Katmada, A., Tsiatsos, T.: Impact of online flexible games on students' attitude towards mathematics. Educ. Technol. Res. Dev. **65**(6), 1451–1470 (2017). https://doi.org/10.1007/s11423-017-9522-5
Mobley, C., Vagias, W.M., DeWard, S.L.: Exploring additional determinants of environmentally responsible behavior: the influence of environmental literature and environmental attitudes. Environ. Behav. **42**(4), 420–447 (2010)
Moghavvemi, S., Paramanathan, T., Rahin, N.M., Sharabati, M.: Student's perceptions towards using e-learning via Facebook. Behav. Inf. Technol. **36**(10), 1081–1100 (2017)

Morgan, M.: From production line to drama school: higher education for the future of tourism. Int. J. Contemp. Hosp. Manag.(2004)

Paris, C.M.: Social constructivism and tourism education. J. Hosp. Leis. Sport. Tour. Educ. (Pre-2012) **10**(2), 103 (2011)

Phipps, R.A., Merisotis, J.P.: What's the difference?: A review of contemporary research on the effectiveness of distance learning in higher education: Institute for Higher Education Policy (1999)

Raptis, G.E., Fidas, C., Avouris, N.: Effects of mixed-reality on players' behaviour and immersion in a cultural tourism game: a cognitive processing perspective. Int. J. Hum Comput Stud. **114**, 69–79 (2018)

Sadeghi, M.: A shift from classroom to distance learning: advantages and limitations. Int. J. Res. Engl. Educ. **4**(1), 80–88 (2019)

Saulnier, B.M., Landry, J.P., Longenecker Jr., H.E., Wagner, T.A.: From teaching to learning: learner-centered teaching and assessment in information systems education. J. Inf. Syst. Educ. **19**(2), 169 (2008)

Schapire, R.E.: The boosting approach to machine learning: an overview. In: Denison, D.D., Hansen, M.H., Holmes, C.C., Mallick, B., Yu, B. (eds.) Nonlinear Estimation and Classification. LNS, vol. 171, pp. 149–171. Springer, New York (2003). https://doi.org/10.1007/978-0-387-21579-2_9

Schott, C.: Virtual fieldtrips and climate change education for tourism students. J. Hosp. Leis. Sport. Tour. Educ. **21**, 13–22 (2017)

Scotland, M.: Higher education program curricula models in tourism and hospitality education: a review of the literature. Online Submission (2006)

Semley, N., Huang, R., Dalton, J.: Feedback for learning development: tourism students' perspective. J. Hosp. Leis. Sport. Tour. Educ. **19**, 41–53 (2016)

Sigala, M.: The evolution of internet pedagogy: Benefits for tourism and hospitality education. J. Hosp. Leis. Sport. Tour. Educ. **1**(2), 29–45 (2002)

Sigala, M.: Investigating the factors determining e-learning effectiveness in tourism and hospitality education. J. Hosp. Tour. Educ. **16**(2), 11–21 (2004)

Sisson, L.G., Adams, A.R.: Essential hospitality management competencies: the importance of soft skills. J. Hosp. Tour. Educ. **25**(3), 131–145 (2013)

Ting, D.H., Cheng, C.F.C.: Developing pro-environmental behaviour: ecotourism fieldtrip and experiences. Int. J. Sustain. High. Educ. (2017)

Tribe, J.: Tourism: a critical business. J. Travel Res. **46**(3), 245–255 (2008)

Tsai, C.-C.: The development of epistemic relativism versus social relativism via online peer assessment, and their relations with epistemological beliefs and internet self-efficacy. J. Educ. Technol. Soc. **15**(2), 309–316 (2012)

Watson, F.F., Bishop, M.C., Ferdinand-James, D.: Instructional strategies to help online students learn: feedback from online students. TechTrends **61**(5), 420–427 (2017)

Webster, R.: Declarative knowledge acquisition in immersive virtual learning environments. Interact. Learn. Environ. **24**(6), 1319–1333 (2016)

Weibel, D., Stricker, D., Wissmath, B.: The use of a virtual learning centre in the context of a university lecture: factors influencing satisfaction and performance. Interact. Learn. Environ. **20**(1), 77–87 (2012)

Xu, F., Buhalis, D., Weber, J.: Serious games and the gamification of tourism. Tour. Manag. **60**, 244–256 (2017)

Yan, Z.: Unprecedented pandemic, unprecedented shift, and unprecedented opportunity. Hum. Behav. Emerg. Technol. **2**(2), 110–112 (2020)

Yang, Y., Zhang, H., Chen, X.: Coronavirus pandemic and tourism: dynamic stochastic general equilibrium modeling of infectious disease outbreak. Ann. Tour. Res. (2020)

Yukselturk, E., Yildirim, Z.: Investigation of interaction, online support, course structure and flexibility as the contributing factors to students' satisfaction in an online certificate program. J. Educ. Technol. Soc. **11**(4), 51–65 (2008)

Zagonari, F.: Balancing tourism education and training. Int. J. Hosp. Manag. **28**(1), 2–9 (2009)

Zehrer, A., Mössenlechner, C.: Industry relations and curricula design in Austrian tourism master programs: a comparative analysis. J. Teach. Travel. Tour. **8**(1), 73–95 (2008)

The Impacts of Digital Note-Taking on Classroom Instruction: A Literature Review

Harrison Hao Yang[1,2](✉), Yinghui Shi[1], Huiyun Yang[1], and Qiuyu Pu[1]

[1] Central China Normal University, Wuhan, China
[2] State University of New York at Oswego, New York, USA
harrison.yang@oswego.edu

Abstract. As an essential part of classroom activities, note-taking can benefit students in learning the materials. Recently, the widespread use of electronic devices, such as laptops, tablet PCs, and smartphones in classroom instruction has gradually moved away from traditional notes such as pen and paper to digital note-taking based on portable electronic devices. Although note-taking with digital devices has been widely used in education, there has been a lack of systematic review on digital note-taking studies. The present study made a comprehensive analysis of relevant studies from four aspects, including theoretical studies, technical studies, studies on users, and empirical studies. Discussion and conclusion are also provided in this paper.

Keywords: Digital note-taking · Classroom instruction · Theoretical studies · Technical studies · Empirical studies

1 Introduction

Note-taking is a common and complex activity that involving knowledge comprehension and collection, and a written production process in classroom instruction. This well-known practice should be understood for both theoretical and practical reasons (Piolat et al. 2005). Note-taking is usually realized under severe time pressure, it is necessary to shorten and reduce the information to take notes quickly (Piolat et al. 2005). With the rapid development of digital technology, the widespread use of electronic devices, such as laptops and tablet PCs, has replaced the role of the traditional devices (e.g., pen, paper) to take notes in the classroom (Kim et al. 2009).

According to a definition provided by Anderson-Inman and Homey (2007), digital note-taking is "a resource that provides markers or note-taking tools in text for later retrieval and uses for studying or completing assignments" (p. 154). Compared with traditional hand-written note-taking, digital note-taking is more malleable and ready to edit, copy, search, organize, share, translate, and repurpose to support reflection, recall, synthesis, and collaboration (Willett et al. 2015). For many students, the use of electronic devices can improve their note-taking speed. From the standpoint of external storage, the ability to take more notes provide students with obvious benefits. The more notes that students recorded, the greater the impact they may have on learning (Bui et al. 2013).

L.-K. Lee et al. (Eds.): ICTE 2020, CCIS 1302, pp. 61–72, 2020.
https://doi.org/10.1007/978-981-33-4594-2_6

Due to its numerous advantages, digital note-taking has been rapidly popularized in various schools. Therefore, digital note-taking has received extensive attention around the world and has become one of the research hotspots of researchers and educators worldwide. While a number of studies regarding the use of digital note-taking have been conducted in recent years, several questions have arisen from this growth. What is the overall impact of digital note-taking in the field of education? What are the focal aspects of digital note-taking that researchers and practitioners have been concerned with? Where has progress been made in digital note-taking? To shed light on these questions and have a better understanding of digital note-taking on teaching and learning, this study intends to provide an overview of the status of digital note-taking in the educational context in recent years. Specifically, the purposes of this study were as follows:

1. To determine what progress has been made in the theoretical and technical aspects regarding the use of digital note-taking in the educational context.
2. To examine the perceptions and attitudes of users toward digital note-taking.
3. To quantify the overall effectiveness of the use of digital note-taking on student learning outcomes in educational experimental studies.

2 Method

To have a better understanding of the research status of digital note-taking in the last two decades, a literature review on digital note-taking was conducted to identify related studies on digital note-taking.

2.1 Data Collection

The Education Resources Information Centre, Elsevier Science Direct online, ProQuest, Taylor & Francis Online, and Springer Link were selected as sources of data. These databases provide a wide multidisciplinary perspective with a variety of citation indexes, content sources, publications (MacLeod et al. 2019). The search was restricted to articles written in English and published in peer-reviewed journals during the period 1999–2019. The specific search keywords utilized in this study were as follows: "digital note-taking" or "digital note" or "electronic note". The search results were screened and processed manually.

Studies were examined to ascertain whether they met the following selection criteria: (a) the content of the article is related to digital note-taking; (b) written in an English peer-reviewed journal. Two researchers searched the databases using the same set of criteria and independently selected articles that met the criteria. If two researchers select different papers, they need to discuss and resolve their differences. To reach a consensus, a third researcher was also brought to discuss and resolve any differences in the article selection process. The initial electronic searches on the databases yielded a total number of 507 relevant articles. After screening an abstract review of each article, a total of 54 eligible academic journal articles were obtained. In addition, this study also used Google Scholar to supplement articles related to digital note-taking. As a result, 17 extra articles were added. A total of 71 studies related to digital note-taking were finally obtained. The subjects of these studies included biology psychology, chemistry, medicine, mathematics, science, psychology, engineering, physics, and computer science, etc. The education levels of these studies included both K-12 and higher education. The number distribution of these studies in each year is shown in Fig. 1.

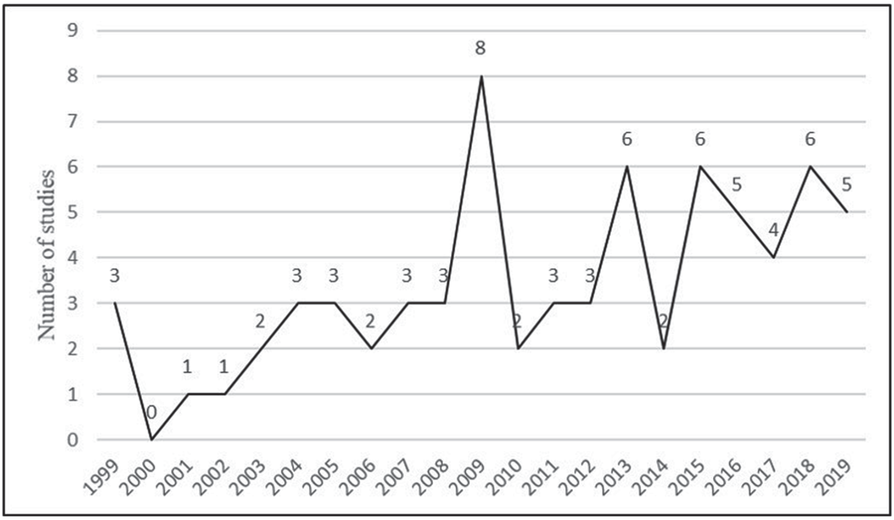


Fig. 1. The quantity distribution of related studies on digital note-taking.

As shown in Fig. 1, research on digital note-taking in the first decade (1999–2008) is at a low level, with studies published not more than 3 per year. However, the average number of studies related to digital note-taking has increased to 5 per year in the second decade (2009–2019). In the last decade, studies on digital note-taking have more than doubled compared to the previous decade. It can be seen that, with the widespread use of digital note-taking in the classroom, the implications of digital note-taking have been of great interest to researchers.

2.2 Studies Classification

Based on the research and practice of digital note-taking in education, this study put the obtained studies into four categories: theoretical studies, technical studies, studies on users, and empirical studies. Among which, theoretical studies mainly include studies related to the instructional process, instructional model, and instructional strategy of digital note-taking. Technical studies mainly include hardware technology and software technology related to digital note-taking. Studies on users refer to studies aiming to examine the perceptions, feelings, and attitudes of the users toward the digital note-taking. Empirical studies mainly include studies with an experimental or quasi-experimental design that investigating the effects of digital note-taking on the outcomes of classroom instruction, such as student learning achievement. If a study belongs to more than one category, then it will be counted into these categories separately. Studies not covered by these four categories are classified as other studies. The research topic categories and the number of studies in each category are shown in Fig. 2.

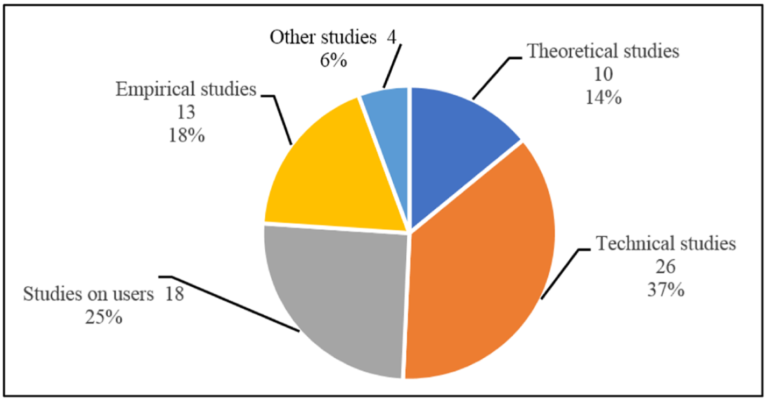


Fig. 2. Thesis research topic distribution map.

As shown in Fig. 2, the number of technical studies and studies on users is relatively larger, which showed that researchers have greater interests in the technical research of digital note-taking, as well as users' perceptions and attitudes toward the application of digital note-taking in the classroom instruction. It seems that improving new products for digital note-taking and thus enhancing users' positive perceptions and attitudes toward digital note-taking has become a research focus. In addition, the quantity of theoretical studies and empirical studies on digital note-taking seems relatively insufficient.

In the next section, we will illustrate thoroughly the research status of digital note-taking from the previously mentioned four categories respectively.

3 Results

3.1 Theoretical Studies on Digital Note-Taking

From the theoretical aspect, studies have been done to introduce some theories and research models to guide the application of digital note-taking in classroom instruction. These theories and models include the multimedia learning theory (MLT) and the model of encoding-review.

The MLT was proposed by Mayer (2001). According to the MLT, learners can obtain more knowledge if both the visual and auditory channels are provided with information. On the other hand, if too much information is delivered, the learning effect will be reduced. Therefore, some researchers have conducted comparative studies of digital note-taking and traditional note-taking from the perspective of note types. For example, Mueller and Oppenheimer (2014) found that taking notes on computers was verbatim, with fewer symbols, charts, and images recorded. Luo et al. (2018) reported that compared with handwritten notes, drawing pictures, charts, and graphs on notebook computers were much slower and more difficult. However, on the whole, traditional note-taking can only act on visual channels, while digital note-taking can act on visual and auditory channels. In addition, Horney et al. (2009) examined the influence of different types of digital note-taking on learning, whether it is typed text (text notes) or recorded voice (voice notes). They found that the voice notes group outperformed their text notes peers on learning science texts. Sencibaugh (2007) found that auditory- and language-dependent strategies (e.g., summarization and finding the main idea) had a greater impact on reading comprehension than visually-dependent strategies (e.g., illustrations in the text and semantic organizers). Together, these findings indicate that voice notes may be a more effective note-taking strategy for students with disabilities and general students (Horney et al. 2009).

DiVesta and Gray (1972) proposed a model of encoding-review and distinguished between two possible functions of note-taking—storage and encoding. Their views are further supported by other researchers. They think that taking notes is to code the content presented in class, which can promote students' understanding of knowledge and storage of knowledge (Peper and Mayer 1978), and the use of notes after class or before the examination can effectively help students improve their homework and test scores (Kiewra 1989). With the popularization of digital note-taking, researchers tend to compare the application of digital note-taking and traditional handwritten notes in classroom instruction based on the model of encoding-review. For instance, Patterson and Patterson (2017) found that the process of traditional handwritten notes was considered to be closely related to coding, so traditional note-taking was generally considered to have better recall ability than notes on computers or other devices. However, Bui et al. (2013) found that taking notes with notebook computers had more advantages than handwritten note-taking, because digital note-taking could quickly, and easily store, search, access, and share information (Walsh and Cho 2013).

3.2 Technical Studies

Numerous technological advancements have triggered an information explosion, and partly changed the way of information management and instructional practices. Advancements in technology such as tablet PCs, mobile applications (apps), and recorded lectures are changing classroom dynamics that affect the way students write and review class notes (Stacy and Cain 2015). Since note-taking has varied purposes in different situations, the effectiveness of various technologies will also depend on the situation (Mueller and Oppenheimer 2016). There are already lots of devices for digital note-taking technology, such as digital pens, digital reading desks, and smartpens. A number of digital note-taking apps and systems have also been developed, in which voice recognition, automatic capture, and other technologies can help students edit notes more easily. Therefore, technical studies can be divided into two parts: hardware and software.

Hardware. Digital note-taking is the use of cell phones, computers, and other digital technology devices to take notes (Jackson 2015). In the process of classroom instruction, using mobile phones to take notes is a common practice for most university students (Mfaume et al. 2018). Reimer et al. (2009) claimed that the popularity of mobile phone technology has made students rely heavily on smartphones when completing some academic tasks. Kim et al. (2009) found that while the technical support for writing with pens on the electronic surface was very advanced, students responded very differently to note-taking devices and that their existing activities were not always well embraced.

Researchers have investigated different digital note-taking systems and corresponding accessories on the effectiveness of digital note-taking. For example, Migos et al. (2016) proposed systems, techniques, and methods for creating digital note-taking cards and provided graphical user interfaces for interacting with digital note-taking cards. Pearson et al. (2011) introduced a document reading interface: a digital reading desk, which enhanced the existing digital reading interaction by adopting effective paper interaction elements and combining these elements with digital enhancement functions. In addition, Miura et al. (2005) developed the AirTransNote system, which used digital pens to record notes on regular paper, saved students the trouble of taking notes with a PC.

The multifunctional digital pen that used for digital note-taking is another research hotspot in technical studies. The digital pen has a variety of functions, allowing students to record information in multiple modes and to view the information as needed. For instance, WizCom Tech has developed reading pens with scanning, reading, and translation functions, a variety of portable scanning pens, and smartpens for assisted learning. These digital pens are reasonable in price, easy to carry and use, and can be used as effective learning aids for various subjects (Shaffer and Schwebach 2015). Moreover, smart pens developed by LivescribeTM can be used to support class notes, provide multimodal tools for student assessment, create reading materials, and support mathematical calculation (Ok and Rao 2017).

With the help of various digital devices, the way students take notes is gradually becoming digital. The development of ICTs enables digital note-taking to effectively help students organize information, meet the various needs of students, and also reduce the cognitive burden of students (Belson et al. 2013).

Software. Although digital notes are easier to read, search and edit, a portion of students are still taking class notes with traditional pen and paper, partly due to the lack of software to supports digital note-taking (Ward and Tatsukawa 2003). Mosleh et al. (2016) conducted an extensive survey and systematic analysis of current digital note-taking software. They found that digital note-taking could not be used to replace traditional note-taking, owing to the problems of complexity, technical learning difficulties, integrity, and low efficiency in the development and implementation of digital note-taking applications.

Nevertheless, several studies regarding the supporting systems for digital note-taking have been conducted to enhance users' experience on digital note-taking. For instance, Xie et al. (2011) proposed a new computer-aided digital note-taking system named D-Note, which could minimize the damage or dirtiness of books. In addition, Chiu et al. (2013) adopted the digital note-taking platform Google Docs, which was a free web-based office suite and data storage service that provides basic word processing functions, such as file creation, editing, and saving. Students could also use the sharing function to share or edit notes with their peers.

Note-taking app has attracted a great deal of attention from researchers. Note-taking apps can record text, image, voice, video, and other multimedia, and have a variety of interface formats, which can be installed in different systems such as PCs, mobile phones, etc. (Mfaume et al. 2018). Researchers have also addressed that most of the note-taking apps are comprehensive knowledge management tools that integrate diary, note, thought map, inspiration record, reading notes, to-do list, etc., so they meet the different needs of most users, but also has more functions and complex structure (Mosleh et al. 2016). Most digital note-taking apps are cloud-based, allowing users to take notes on a tablet and real-time updates of content after logging in the account at any terminal (Orndorff III 2015).

Overall, mobile technology and note-taking apps remove time and space constraints and become an efficient way to communicate, collaborate, and share knowledge (Sheng et al. 2010; Willett et al. 2015). Digital note-taking technologies can not only record new information in a timely, complete, and precise manner, but can also quickly scan, organize, and modify existing information (Walsh and Cho 2013).

3.3 Studies on Users

The development of digital note-taking is influencing people's daily life imperceptibly. It can help users to store, manage, and reuse knowledge in a large capacity, fast, and shared way (Willett et al. 2015). Concerning record information by digital note-taking, recording devices like PCs, mobile phones, and other electronic communication devices can be used to record, classify and share information needed by users in a timely manner (Mfaume et al. 2018). Users can choose appropriate note-taking functions according to their hands-on digital devices, which provides great convenience for users to take notes in the process of learning.

During the process of digital note-taking, the users may have different experiences, and their perceptions and attitudes toward digital note-taking are thus varied. Several studies investigate users' acceptance, attitudes, and perceptions toward the use of digital note-taking. For example, Palaigeorgiou et al. (2006) identified obstacles for students' acceptance of using eVernotes, an application for facilitating verbatim note-taking and enabling the creation of multiple notes' associations. The results showed the overwhelming majority of students provided positive evaluations of the eVernotes, while several participants were unwilling to use the eVerNotes because of their incompatible preconceived expectations and earlier experiences. Amick and Cross (2014) examined college students' use of iPads to take notes in a Chemistry course. The results showed that student feedback about using iPads in class was very positive, and 75% of students preferring to use an iPad to take notes rather than the traditional paper and pencil method. Similarly, Morehead et al. (2019) reported that students highly valued the ability to use PCs to take notes. Bates (2016) argued that digital note-taking could help professors organize and manage notes effectively, because note-taking apps provided an organizational structure that allowed users to create individual notebooks for notes related to certain topics. By doing this, notes were systematically stored and easily retrieved, greatly improving the working efficiency of users.

However, other studies indicate that paper notes are the most direct, fastest, and convenient tool for recording and expressing scattered information. Although keyboard input is faster, it is not as convenient as handwriting in the convenience of using various generative strategies (Fiorella and Mayer 2017). Furthermore, Yang and Lin (2015) found that despite some studies showing handwritten notes could improve understanding and learning outcomes, many students indicated their preference for digital note-taking (Yang and Lin 2015).

3.4 Empirical Studies

The effects of digital note-taking on students learning outcomes have been examined by a number of empirical studies in the past two decades. Researches have shown that compared with traditional note-taking, digital note-taking can improve the learning outcomes of students. For example, Sun and Li (2019) conducted a three-month quasi-experiment on 72 first-year high school students to evaluate how digital notes using mobile terminals affect student performance. Digital notes referenced in the study are a method to write, record, store, and share notes on mobile terminals. The results showed that the scores of students who took digital notes were significantly higher than those of students who took notes with the traditional method, and digital notes could improve the learning outcomes in all three categories of students programming knowledge. Luo et al. (2018) explored how the process and product functions of laptop and longhand note-taking differentially impact student's notes and achievement. The results indicated that the process of taking notes was more advantageous than the product function of notes (i.e., better image-related learning and similar text-related learning) when taking notes on a laptop. When making longhand notes, the product function of the note was more beneficial than the process function of the note (i.e., better text-related learning and similar image-related learning).

In addition, digital note-taking may be used as a more effective note-taking strategy for students in special education. Teachers can pay attention to students with disabilities who often have difficulty writing text notes. For instance, Horney et al. (2009) investigated the effects of text notes and voice notes of fifth-grade students on comprehension of science texts. The results showed that students in special education could use voice notes instead of text notes to achieve greater test scores. Moreover, Patti and Garland (2015) found that reading pens helped students with reading difficulties improve reading accuracy and comprehension.

However, although digital note-taking has great advantages over traditional note-taking, in some respects, digital note-taking may not be more conducive to student records than traditional note-taking. For example, in the study of Mueller and Oppenheimer (2014), three experiments were conducted to investigate whether taking note-taking on a laptop versus writing longhand affects academic performance. In all three experiments, the laptop note-taking contained more words and more verbatim lecture strings of recorded notes than the longhand note-taking. But to a point, if the notes were taken indiscriminately or by mindlessly transcribing content, the benefit disappeared. Moreover, longhand note-takers might have to do more processing than laptop note-takers, so selecting more important information to include in their notes, which enabled them to study this content more effectively. Bretzing and Kulhavy (1979) pointed out that verbatim note-taking was more problematic for conceptual items. Therefore, on conceptual questions, all three studies by Mueller and Oppenheimer (2014) found that students who took note-taking on laptops performed worse than students who took notes longhand.

4 Discussion and Conclusion

With the development of ICTs, there is an increasing number of users begun to use digital note-taking instead of traditional note-taking. Accordingly, a large number of studies have been conducted to reflect this boom of digital note-taking in classroom instruction. This study presented a literature review on eligible studies in four categories. The main contents of digital note-taking were shown in the third section. The results showed that (1) several theories and research models had been employed to guide the implementation of digital note-taking in classroom instruction; (2) different kinds of digital devices (hardware) and a variety of systems and apps (software) for digital note-taking had been developed to support users' note-taking; (3) users held a positive perception and attitude towards digital note-taking generally, but there had been some obstacles (e.g., did not achieve the expected results or due to economic problems, etc.); and (4) there was a mixed result on the effectiveness of digital note-taking on student learning outcomes. A possible explanation for this might be that immature software and hardware tools hinder the use of digital note-taking (Fox 2005), or students have not selected the appropriate digital note-taking devices for classroom activities.

There is no doubt that digital note-taking has many advantages compared with the traditional note-taking method. For example, digital notes can be quickly stored, searched, accessed, and shared. Digital note-taking can also help teachers organize and manage notes more effectively. However, someone believed that the use of digital devices providing students with opportunities to distract their attention. Therefore, when using digital note-taking in classroom instruction, teachers and students should pay attention to this issue and choose the appropriate digital note-taking tool. For example, 1) selecting a digital pen that is at a reasonable price, easy to carry and use, and suitable to subjects (Shaffer and Schwebach 2015); 2) selecting digital note-taking applications that can record text, image, voice, video, and other multimedia (Chiu et al. 2013). Only in this way can the student's learning outcomes be improved to a greater extent under the conditions of improving student participation and not distracting the student's attention.

Despite the promising results from the previous research, some studies point out that taking notes with digital devices is much slower than that of traditional pen and paper, and users' practices are not always well supported (e.g., Kim et al. 2009). It is still less clear how different pedagogical approaches in classroom teaching have an impact on the effect and demand of students' note-taking. To this end, future studies are suggested to develop more advanced digital devices with more useful features to facilitate users' digital note-taking, such as classroom recording, and to explore more effective pedagogical approaches, to meet the needs of users' digital note-taking in different situations and thus improving users' experiences with digital note-taking. Moreover, since the proportion of empirical studies is relatively low, future studies are encouraged to pay more attention to the effectiveness of digital note-taking on students' learning outcomes.

Acknowledgment. This study was supported by the Ministry of Education of the Humanities and Social Science project (Project No.: 18YJC880074).

References

Amick, A.W., Cross, N.: An almost paperless organic chemistry course with the use of iPads. J. Chem. Educ. **91**(5), 753–756 (2014)

Anderson-Inman, L., Homey, M.A.: Supported eText: assistive technology through text transformations. Read. Res. Q. **42**(1), 153–160 (2007)

Bates, C.C.: Using note-taking apps to support professional development. Kappa Delta Pi Rec. **52**(2), 86–90 (2016)

Belson, S.I., Hartmann, D., Sherman, J.: Digital note taking: the use of electronic pens with students with specific learning disabilities. J. Spec. Educ. Technol. **28**(2), 13–24 (2013)

Bretzing, B.H., Kulhavy, R.W.: Notetaking and depth of processing. Contemp. Educ. Psychol. **4**, 145–153 (1979)

Bui, D.C., Myerson, J., Hale, S.: Note-taking with computers: exploring alternative strategies for improved recall. J. Educ. Psychol. **105**(2), 299–309 (2013)

Chiu, C.H., Wu, C.Y., Cheng, H.W.: Integrating reviewing strategies into shared electronic note-taking: questioning, summarizing and note reading. Comput. Educ. **67**, 229–238 (2013)

DiVesta, F.J.D., Gray, G.S.: Listening and note taking. J. Educ. Psychol. **63**(1), 8–14 (1972)

Fiorella, L., Mayer, R.E.: Spontaneous spatial strategy use in learning from scientific text. Contemp. Educ. Psychol. **49**, 66–79 (2017)

Fox, J.R.: A survey of electronic note-taking behavior in information and library science student. Master's thesis, School of information and Library Science, University of North Carolina at Chapel Hill (2005)

Horney, M.A., Anderson-Inman, L., Terrazas-Arellanes, F., Schulte, W., Mundorf, J., Wiseman, S., et al.: Exploring the effects of digital note taking on student comprehension of science texts. J. Spec. Educ. Technol. **24**(3), 45–61 (2009)

Jackson, E.: Comparing analog and digital note-taking among middle school students. Doctor's theses, School of education, Northcentral University (2015)

Kiewra, K.A.: A review of note-taking: the encoding-storage paradigm and beyond. Educ. Psychol. Rev. **1**(2), 147–172 (1989)

Kim, K., Turner, S., Pérez-Quiñones, M.A.: Requirements for electronic note taking systems: a field study of note taking in university classrooms. Educ. Inf. Technol. **14**(3), 255–283 (2009)

Luo, L.L., Kiewra, K.A., Flanigan, A.E., Peteranetz, M.S.: Laptop versus longhand note taking: effects on lecture notes and achievement. Instr. Sci. **46**(6), 947–971 (2018)

MacLeod, J., Yang, H.H., Shi, Y.H.: Student-to-student connectedness in higher education: a systematic literature review. J. Comput. High. Educ. **31**(2), 426–448 (2019)

Mayer, R.E.: Multimedia Learning. Cambridge University Press, Cambridge (2001)

Mfaume, H., Bilinga, M., Mgaya, R.: From paper and pencil to mobile phone photo note-taking among Tanzanian university students: extent, motives and impact on learning. Int. J. Educ. Dev. Using Inf. Commun. Technol. **14**(2), 83–98 (2018)

Migos, C.J., Capela, J.C., Hagele, M., Maclean, A.C., Emanuel, M.F., Murrett, M., et al.: Creating and viewing digital note cards, US (2016)

Miura, M., Kunifuji, S., Shizuki, B., Tanaka, J.: AirTransNote: augmented classrooms with digital pen devices and RFID tags. In: IEEE International Workshop on Wireless & Mobile Technologies in Education. IEEE (2005)

Morehead, K., Dunlosky, J., Rawson, K.A., Blasiman, R., Hollis, R.B.: Note-taking habits of 21st century college students: implications for student learning, memory, and achievement. Memory **27**(6), 807–819 (2019)

Mosleh, M.A.A., Baba, M.S., Malek, S., Alhussein, M.A.: Challenges of digital note taking. In: Sulaiman, H.A., Othman, M.A., Othman, M.F.I., Rahim, Y.A., Pee, N.C. (eds.) Advanced Computer and Communication Engineering Technology. LNEE, vol. 362, pp. 211–231. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-24584-3_19

Mueller, P.A., Oppenheimer, D.M.: The pen is mightier than the keyboard: advantages of longhand over laptop note taking. Psychol. Sci. **25**(6), 1159–1168 (2014)

Mueller, P.A., Oppenheimer, D.M.: Technology and note-taking in the classroom, boardroom, hospital room, and courtroom. Trends Neurosci. Educ. **5**(3), 139–145 (2016)

Ok, M.W., Rao, K.: Using a digital pen to support secondary students with learning disabilities. Interv. Sch. Clin. **53**(1), 36–43 (2017)

Orndorff III, H.N.: Collaborative note-taking: the impact of cloud computing on classroom performance. Int. J. Teach. Learn. High. Educ. **27**(3), 340–351 (2015)

Palaigeorgiou, G.E., Despotakis, T.D., Demetriadis, S., Tsoukalas, I.A.: Synergies and barriers with electronic verbatim notes (eVerNotes): note taking and report writing with eVerNotes. J. Comput. Assist. Learn. **22**(1), 74–85 (2006)

Patterson, R.W., Patterson, R.M.: Computers and productivity: evidence from laptop use in the college classroom. Econ. Educ. Rev. **57**, 66–79 (2017)

Patti, A.L., Garland, K.V.: Smartpen applications for meeting the needs of students with learning disabilities in inclusive classrooms. J. Spec. Educ. Technol. **30**(4), 238–244 (2015)

Pearson, J., Buchanan, G., Thimbleby, H.: The reading desk: supporting lightweight note-taking in digital documents. In: Gradmann, S., Borri, F., Meghini, C., Schuldt, H. (eds.) TPDL 2011. LNCS, vol. 6966, pp. 438–441. Springer, Heidelberg (2011). https://doi.org/10.1007/978-3-642-24469-8_44

Peper, R.J., Mayer, R.E.: Note taking as a generative activity. J. Educ. Psychol. **70**(4), 514–522 (1978)

Piolat, A., Olive, T., Kellogg, R.T.: Cognitive effort during note taking. Appl. Cogn. Psychol. **19**(3), 291–312 (2005)

Reimer, Y., Brimhall, E., Cao, C., O'Reilly, K.: Empirical user studies inform the design of an e-note taking and information assimilation system for students in higher education. J. Comput. Educ. **52**(4), 893–913 (2009)

Sencibaugh, J.: Meta-analysis of reading comprehension interventions for students with learning disabilities: Strategies and implications. Read. Improv. **44**, 6–22 (2007)

Shaffer, A.K., Schwebach, J.R.: Usefulness of Livescribe web recordings as supplemental resources for a large lecture undergraduate course. J. Coll. Sci. Teach. **44**(4), 54–60 (2015)

Sheng, H., Siau, K., Nah, F.F.: Understanding the values of mobile technology in education: A value-focused thinking approach. ACM SIGMIS Database **41**(2), 25–44 (2010)

Stacy, E.M., Cain, J.: Note-taking and handouts in the digital age. Am. J. Pharm. Educ. **79**(7), 107 (2015)

Sun, D., Li, Y.: Effectiveness of digital note-taking on students' performance in declarative, procedural and conditional knowledge learning. Int. J. Emerg. Technol. Learn. **14**(18), 108–119 (2019)

Walsh, E., Cho, I.: Using Evernote as an electronic lab notebook in a translational science laboratory. J. Lab. Autom. **18**(3), 229–234 (2013)

Ward, N., Tatsukawa, H.: A tool for taking class notes. Int. J. Hum Comput Stud. **59**(6), 959–981 (2003)

Willett, W., Goffin, P., Isenberg, P.: Understanding digital note-taking practice for visualization. IEEE Comput. Graph. Appl. **35**(4), 38–51 (2015)

Xie, D., Zhou, Z., Wu, J.: D-Note: computer-aided digital note taking system on physical book. In: Sixth International Conference on Image & Graphics. IEEE Computer Society (2011)
Yang, Y.F., Lin, Y.Y.: Online collaborative note-taking strategies to foster EFL beginners' literacy development. System **52**, 127–138 (2015)

Learning Analysis and Assessment

An Empirical Study of the Effect of ASR-Supported English Reading Aloud Practices on Pronunciation Accuracy

Yuting Yuan and Xiaobin Liu(✉)

South China Normal University, Guangzhou, China
liuxiaobin@m.scnu.edu.cn

Abstract. Automatic Speech Recognition (ASR) technology has excellent potential for improving learners' English pronunciation. However, the specific effects of ASR on pronunciation improvement are still unclear. Taking learners' language proficiency into account, the present study aims to investigate the specific effects of ASR-aided reading aloud practices on English pronunciation accuracy. 50 senior high school students from a city in southern China are participants. They are divided into two groups: low- and high-proficient groups. Two tests and specific pronunciation error analysis are used to test students' pronunciation improvement, while a post-study survey is conducted to explore students' perception of using ASR. The results indicate that, for low-proficient learners, ASR-supported reading aloud practices help them improve global pronunciation accuracy and correct word error and deletion error; while for high-proficient learners, the training doesn't improve their global pronunciation but helps them correct deletion error and addition error. Both groups show positive perception of ASR-supported training.

Keywords: ASR · English reading aloud practice · Pronunciation accuracy · Different language proficiency

1 Introduction

Pronunciation is the basis of language learning, which plays a key role in the development of language competence. Intelligible pronunciation not only can improve learners' comprehensibility in oral communication, but also can improve their speaking fluency (Morley 1991; Howlader 1984). However, pronunciation has been frequently ignored in foreign language (FL) research and teaching (Derwing and Munro 2005). In China, most teachers put much emphasis on other language items and skills like vocabulary, grammar, reading and writing, while pronunciation teaching only focuses on the recognition of English phonetic symbols (Luo and Zhang 2002). This approach cannot effectively and efficiently improve learners' pronunciation. According to Xu and Zeng (2015), there are many pronunciation problems like substitution, addition, deletion, mispronunciation, error in words or lexical chunks and word stress error. Middle and high school is the critical period to acquire pronunciation (Wang 2010), hence, it is of great necessity to explore some novel effective approaches for high school students to practice pronunciation.

L.-K. Lee et al. (Eds.): ICTE 2020, CCIS 1302, pp. 75–87, 2020.
https://doi.org/10.1007/978-981-33-4594-2_7

Fortunately, with increasing development of information technology, there is a variety of speech tools and technologies that can help FL learners work on their pronunciation. ASR is one of the promising technologies. There are plenty of intelligent ASR-based applications such as *Tell Me More* (Elimat and AbuSeileek 2014)*, Rosetta Stone* (Sharifi et al. 2015)*, My English Tutor* (Tsai 2003), which provide novel methods and resources for learners to practice pronunciation. Quite a few studies have proved that ASR can facilitate learners to improve their pronunciation (Liu et al. 2019; Elimat and AbuSeileek 2014). However, these researches mostly focus on the effect of ASR on the global pronunciation improvement, without involving specific effects of ASR, i.e., whether ASR can correct every pronunciation error and has the same effect on learners of different proficiency.

Therefore, combining ASR-based app *Oral English Drill & Test* with reading aloud activity, the present study aims to study the specific effects of ASR-aided reading aloud practices on English pronunciation accuracy in computer-based test, hoping to explore an effective way to deal with learners' common pronunciation problems.

2 Previous Studies of ASR-Aided Pronunciation Teaching and Learning

In recent years, computer-assisted pronunciation teaching (CAPT) and mobile-assisted language learning (MALL) tools and applications give many possibilities of FL pronunciation teaching and learning. As an independent, machine-based process of decoding and transcribing oral speech into texts (Levis and Suvorov 2014), ASR is supposed to be helpful in teaching and learning pronunciation if well applied (Eskenazi 1999a; Neri et al. 2003). ASR can bring various advantages to pronunciation training. For example, it provides a large quantity of audio records of native speakers and oral output practices; its automatic feedback enables learners to realize their pronunciation problems and correct them in time, which can make up for the deficiency of traditional pronunciation teaching; whenever and wherever possible, learners can independently use ASR-based application to practice pronunciation, which may also reduce learners' pressure and anxiety (Ehsani and Knodt 1998; Eskenazi 1999b; Neri et al. 2001).

Many researchers conducted studies to further investigate the effectiveness of ASR-aided pronunciation teaching and learning. The effect of ASR on the improvement of pronunciation performance was a main focus. Neri et al. (2008) studied the feasibility and effectiveness of ASR-based CAPT system *Dutch-CAPT* on improving the target phonemes pronunciation. The results showed that a system with automatic feedback not only helped learners improve segmental quality, but also speeded up their segmental improvement. In Elimat and AbuSeileek's study (2014), 64 third-grade students used ASR-based application *Tell Me More Performance English* to learn pronunciation of word, sentence, and real-life dialogue through individual work, pair work, or group work, and the result indicated that ASR-aided pronunciation training was more effective than traditional approach and individual work was better than the others. Liakin et al. (2014) pointed out that mobile ASR-based pronunciation instruction significantly improved learners' pronunciation production of French /y/ but cannot significantly improve its pronunciation perception. Then, Liakin et al. (2017) found out that learners recognized the pedagogical use of TTS and ASR tools and enjoyed using them. Liu et al. (2018) found out that automatic speech evaluation (ASE) app *Fluent English* can improve pronunciation and intonation and the learning strategies included repetition, reinforcement, visualization, resourcing, small-pacing, task-taking and self-encouragement. Liu et al. (2019) proved that ASR-based application *iFlytek Voice Input* can assist Chinese advanced EFL learners to correct some typical Chinese pronunciation errors.

The relationship between ASR and pronunciation autonomy also attracted some scholars' attention. Kruk (2012) asked 45 Polish senior high school learners to use ASR-based online resources to practice pronunciation. The results revealed that the majority of subjects showed more learning autonomy and outperformed their counterparts on tests. McCrocklin (2016) found out that ASR-aided approach can increase learners' beliefs of autonomy and the feedback of ASR enabled learners to practice pronunciation more autonomously.

Learners' perception and attitude of using ASR was another consideration. From the perspective of learning styles, Hsu (2015) discovered that visual style and kinaesthetic style were significantly influential to perceived ease of use and this significantly led to perceived usefulness which determined learners' attitudes to continue to use ASR-based CAPT. Tsai (2019) focused on the difficulties encountered by learners and found out that collaborative work had better effects like monitoring and modifying each other's speech production as well as enhancing learning motivation.

Studies cited above prove that ASR can help learners improve English pronunciation, which lays some foundation for the ASR-supported approach in pronunciation training that the present study attempts to explore. What sets it apart from previous studies is its concentration on specific effects of ASR on learners of different proficiency which aims to reduce Chinese EFL learners' common pronunciation problems.

3 Research Design

3.1 Research Questions

The present study tries to answer the following three questions:

RQ1. Do ASR-supported reading aloud practices have different effects on pronunciation improvement of different language proficiency?
RQ2. For learners of different language proficiency, what pronunciation errors can ASR-supported reading aloud practices help correct?
RQ3. What are learners' perceptions of ASR-supported reading aloud practices?

To answer RQ1 and RQ2, a 28-day experiment of ASR-supported reading aloud practices was conducted. To answer RQ3, a post-study questionnaire survey was carried out.

3.2 Participants

The training experiment took place from January to February of 2020. 97 grade-three students in a senior high school in Lianzhou, Guangdong, China participated in the 28-day ASR-supported reading aloud training. 50 of them were chosen as participants (mean age = 17.82, SD = 0.63). The selection of participants was according to the assessment of learners' language proficiency based on two scores: score of pronunciation pre-test and score of the final exam in last term which includes reading, vocabulary, grammar and writing. 25 students who got lower scores were classified as the low-proficient group; while 25 students who got higher scores in the two tests were regarded as the high-proficient group. The differences of pronunciation pre-test scores and final exam scores between the two groups are shown in Table 1 and Table 2. Both differences of pronunciation pre-test scores and final exam scores are significant between the two groups.

Table 1. Independent-samples t test of pre-test pronunciation scores

	N	M	SD	t	p (2-tailed)
Low-proficient group	25	12.53	1.65004	−11.568	.000
High-proficient group	25	16.44	.36105		

$p^* < 0.05$

Table 2. Independent-samples t test of final exam scores

	N	M	SD	t	p (2-tailed)
Low-proficient group	25	56.55	15.71978	−7.820	.000
High-proficient group	25	91.50	15.88134		

$p^* < 0.05$

3.3 Instruments

1. Materials: Texts from *New Concept English 2.*

2. ASR-based application: *Oral English Drill & Test* app. It is an oral English training application with ASR technology developed by Associate Professor Liu Xiao-bin from South China Normal University. It can be used for pronunciation self-assessment and correction. After students finishing oral reading, this application uses iFlytek speech recognition to convert speech to text and evaluates the speech by comparing with original script. It contains the recordings and texts of textbooks *New Senior English for China* and *New Concept English 2*. It has two functions: one is imitating and reading aloud; another is free training (Fig. 1). Participants are required to use these two functions to practice texts in *New Concept English 2.*

3. Tests: Reading aloud tasks in 2017 and 2018 CELST in Guangdong, China: students are required to watch a video clip and read after the speaker in the video. These reading aloud tasks have similar difficulty and their mean difficulty coefficient is 0.73.

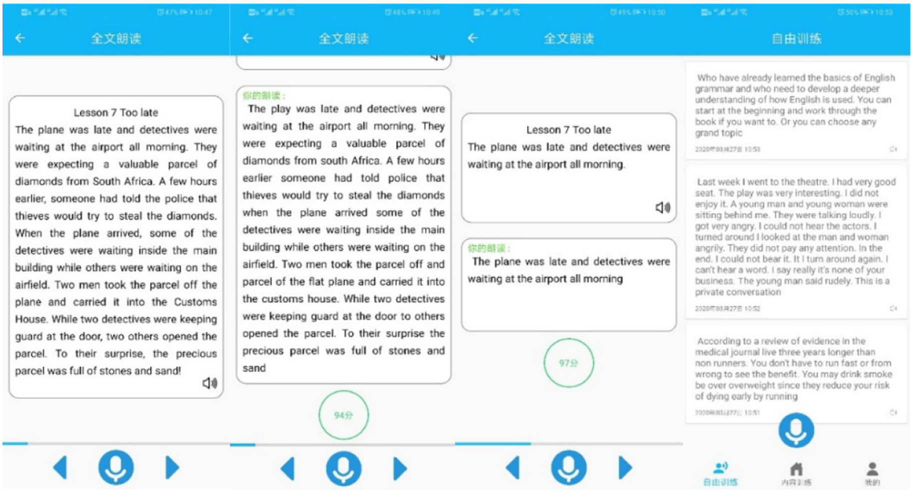


Fig. 1. Screenshot of the main function in *Oral English Drill & Test*

4. Questionnaire: It is adapted from Hsu (2015), including 13 multiple-choice questions about participants' perceptions of ASR-supported reading aloud practices which uses Likert Five Rating Scale (1 = strongly disagree; 2 = disagree; 3 = neutral; 4 = agree; 5 = strongly agree). The reliability of questionnaire was ensured by Cronbach alpha for the internal consistency of items (Cronbach alpha = 0.929).

3.4 Procedure

The procedure includes pre-test, ASR training experiment, post-test and questionnaire survey.

1. Pre-test: It required participants to finish reading aloud task in language laboratory.

2. Experiment: Participants took part in a 28-day pronunciation training. Firstly, participants used imitating and reading aloud function of *Oral English Drill & Test* to imitate and read aloud a text sentence by sentence. They practiced their poor pronunciation again and again according to the score and feedback given by the app. Secondly, participants chose free training function to practice the same text without modelling to test their pronunciation and recorded their speech. Thirdly, participants wrote down those words and sentences they cannot pronounce well and reviewed them regularly. Lastly, participants uploaded their recordings and screenshot of practicing in *Oral English Drill & Test* in online learning group. Everyday participants finished one lesson of *New Concept English 2*. Meanwhile, we also taught some pronunciation knowledge and pointed out common pronunciation errors to further help them correct poor pronunciation.

3. Post-test & Questionnaire survey: At last, participants were required to take part in five sets of reading aloud tasks in a computer testing program *EKING.COM* application and fill out the questionnaire in *wjx.cn*. The testing system in *EKING.COM* is similar to that of language laboratory in school, which has high reliability in listening and speaking test.

3.5 Data Collection and Analysis

Both pre- and post-tests were computer-based tests. The computer testing program recorded participants' reading aloud and graded the recordings according to the scoring criteria of reading aloud task in CELST (with a full score of 20). Specific pronunciation accuracy was analyzed by calculating and classifying four common pronunciation error types (Table 3).

Table 3. Definitions of common pronunciation errors (Isaacs and Trofimovich 2012; Xu and Zeng 2015)

Error types	Definition	Example
Phonemic error	Vowel or consonant is mispronounced.	*cub* spoken as [kɒb]; *very* spoken as [ˈwerɪ]
Addition error	Add sounds that are not in the correct pronunciation.	*ten* spoken as tens; *spread* spoken as [sepred]
Deletion error	Delete sounds in the correct pronunciation.	*cancers* spoken as *cancer*; *hunting* spoken as *hunt*
Word error	(1) A word is mispronounced as another word; (2) The pronunciation of a word cannot be recognized.	*she* spoken as *he*

4 Results and Discussion

4.1 Results of Overall Pronunciation Accuracy

As shown in Table 4, the mean scores of the post-test in low- and high-proficient groups were higher than those of pre-test to varying extent. The mean score increased by 1.03 in low-proficient group and increased by 0.34 in high-proficient group. The differences from pre- to post-test in low-proficient group were statistically significant ($t = -3.996$, $p = 0.001 < 0.05$), while difference in high-proficient group were not ($t = -1.733$, $p = 0.096 > 0.05$). This indicated that ASR-supported reading aloud practices can significantly improve English pronunciation accuracy of low-proficient learners, but for high-proficient learners, the effect on the improvement of pronunciation accuracy seems not so evident.

Table 4. Paired-samples t test of total scores

		N	M	SD	t	p
Low-proficient group	Pre-test	25	12.53	1.65004	−3.996	.001
	Post-test	25	13.56	1.97415		
High-proficient group	Pre-test	25	16.44	.36105	−1.733	.096
	Post-test	25	16.78	.96960		

$p^* < 0.05$

4.2 Results of Specific Pronunciation Accuracy

To find out what pronunciation problems ASR can help solve, specific pronunciation errors of each group at pre- and post-test were analyzed and compared.

For low-proficient group, as the data shown in Table 5, except for phonemic error, the mean numbers of other errors were lower at post-test than those at pre-test. The mean number of phonemic error was slightly higher at post-test than that at pre-test, but there was no significant difference ($t = -1.633$ $p = 0.116 > 0.05$). The mean number of word error and deletion error dropped by 4.24 and 1.32($t = 3.046$, $p = 0.006 < 0.05$; $t = 2.135$, $p = 0.043 < 0.05$), which had significant difference. The mean number of addition error dropped by 0.24, but it showed no significant difference ($t = 1.218$, $p = 0.235 > 0.05$). Thus, it can be concluded that ASR-supported reading aloud practices were effective in helping learners of low proficiency solve word error and deletion error.

Table 5. Paired-samples t test of specific items in low-proficient group

		N	M	SD	t	p
Phonemic error	Pre-test	25	1.56	1.19304	−1.633	.116
	Post-test	25	2.36	2.15793		
Word error	Pre-test	25	13.32	7.58683	3.046	.006
	Post-test	25	9.08	7.18749		
Deletion error	Pre-test	25	3.68	2.19317	2.135	.043
	Post-test	25	2.36	1.70489		
Addition error	Pre-test	25	1.52	1.53080	1.218	.235
	Post-test	25	1.20	1.29099		

$p^* < 0.05$

For high-proficient group, as the data shown in Table 6, the mean numbers of all the four errors were somewhat lower at post-test than those at pre-test. The mean number of phonemic error and word error reduced by 0.36 and 0.52 respectively, which were both not significant ($t = 1.565$, $p = 0.131 > 0.05$; $t = 1.834$, $p = 0.079 > 0.05$). The mean number of deletion error and addition error decreased by 1.32 and 0.6 respectively, both of which showed significant difference ($t = 2.799$, $p = 0.01 < 0.05$; $t = 2.167$, $p = 0.04 < 0.05$).

Table 6. Paired-samples t test of specific items in high-proficient group

		N	M	SD	t	p
Phonemic error	Pre-test	25	1.40	1.19024	1.565	.131
	Post-test	25	1.04	1.13578		
Word error	Pre-test	25	1.80	1.93649	1.834	.079
	Post-test	25	1.28	1.83757		
Deletion error	Pre-test	25	2.56	2.67831	2.799	.010
	Post-test	25	1.24	1.23423		
Addition error	Pre-test	25	1.12	1.30128	2.167	.040
	Post-test	25	.52	.65320		

$p^* < 0.05$

These results revealed that ASR-supported reading aloud practices can help learners of high proficiency correct deletion error and addition error.

4.3 Results of Questionnaire Survey

The results of the questionnaire survey are shown in Table 7 and Fig. 2. The mean overall scores of low-proficient group and high-proficient group were 50 scores above and there was no significant difference between them. The mean scores of each statement in both groups were 3.8 scores above. These results showed that learners had a positive perception of ASR-supported reading aloud practices, regardless of language proficiency.

Table 7. Independent-samples t test of Questionnaire

	Low-proficient group (N = 25)		High-proficient group (N = 25)		t	p (2-tailed)
	M	SD	M	SD		
Score	54.40	6.595	52.84	6.78	.825	.414

$p^* < 0.05$

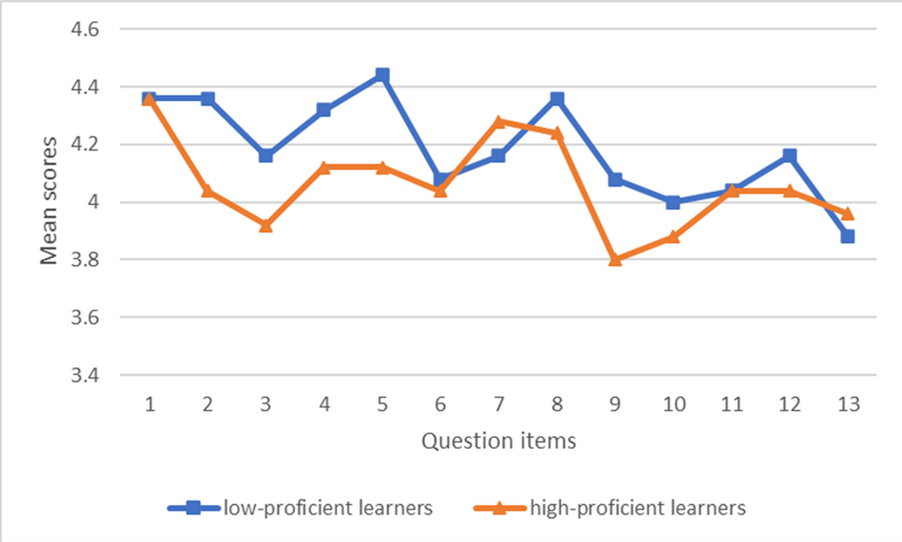


Fig. 2. Participants' responses in questionnaire survey

4.4 Discussion

The goal of the present study is to explore the potential of the ASR-based app *Oral English Drill & Test* to help EFL learners improve their pronunciation accuracy by correcting some common pronunciation errors.

RQ1 aims to testify whether ASR-supported reading aloud practices have different effects on pronunciation improvement of different language proficiency. The above results of total score indicate that ASR-supported reading aloud practices can significantly improve low-proficient learners' pronunciation accuracy, but they have no significant role in improving pronunciation of high-proficient learners. This finding is similar to Hincks' finding (2003) that practice with speech recognition program is beneficial to learners with strong foreign accent but is of limited value for learners with better pronunciation. This may be explained by the fact that ASR transcribes speech into text, which enables learners to notice more obvious errors but these errors are rarely seen in high-proficient group.

Although ASR-supported training cannot significantly improve global pronunciation accuracy of high-proficient learners, it can help them correct some pronunciation errors. As Neri et al. (2008) find out that ASR-based feedback is effective in correcting some pronunciation errors, the results of RQ2 reveal that ASR-supported reading aloud practices can help low-proficient learners correct word error and deletion error and help high-proficient learners correct deletion error and addition error.

Specifically, for phonemic error, its number in both two groups doesn't drop significantly. This can be explained by two aspects. For one thing, although unrecognizable word sounds in low-proficient group are improved, the pronunciation of some words still cannot be perfectly correct, that is some word errors changes into phonemic errors. Actually, word error is more serious than phonemic error, for it is easier to cause failure in intelligibility, therefore, this change can be regarded as a kind of improvement on pronunciation. For another, phonemic error is subtle and many problematic phonemes are similar in pronunciation and some tend to be fossilized, so it may be difficult for learners to correct every subtle phoneme in 28 days without explicit instruction. This is favor of the Pennington's statement (1999) that it is difficult to alter fossilized pronunciation errors without explicit instruction. This may also indicate that ASR has limited value to correct subtle problematic phonemes. For word error, its number reduces significantly in low-proficient group, but not in high-proficient group. This is probably because word error in high-proficient group is not evident at the start. However, in low-proficient group, there are quite a few unrecognizable sounds in their recordings at pre-test, which is difficult to understand their utterance. Words that are frequently mispronounced included polysyllabic words like *phenomenon* and *spectacular*, and those words have similar spelling such as *explode* and *explore*. Fortunately, these phenomena are improved substantially at post-test and almost all words can be recognized. For deletion error, it shows a significant difference in both two groups. Deletion error appeared frequently at pre-test is deleting sounds of inflectional suffix *-ed* and *-s* and swallowing a whole syllable in polysyllabic words. At post-test, its number drops significantly in both two groups. This result reveals that ASR can help learners notice the sounds they often neglected in the past. For addition error, its number reduces significantly in high-proficient group, but not in low-proficient group. Addition error prevalent at pre-test is adding sounds of [t] and [s] at the end of a word, adding schwa [ə] involuntarily, and unintentionally inserting a vowel sound in consonant clusters. These phenomena are also improved at post-test, especially in high-proficient group, which indicates that ASR can help learners broken their bad pronunciation habits of adding sounds. Although it has no significant difference in low-proficient group, we are hopeful that it can be improved if there is longer training length.

We also concern learners' perception of ASR-supported reading aloud practices in RQ3. Because technology-assisted learning differs from traditional learning, it is necessary to know learner's perception and acceptance of it. The potential advantages of CALL-based training cannot be in vain when learners are willing to use it (Hsu 2015). From the results of questionnaire survey, most learners in the two groups accept the ASR-based app and perceive it as a useful tool to help them improve pronunciation. This indicates that ASR-based app here doesn't cause frustration and bring additional cognitive burdens to learners, so it can be fully used to help learners practice pronunciation.

From above discussion, we can conclude that ASR is effective in helping correct more noticeable pronunciation errors like word error, deletion error and addition error. ASR recognizes learners' utterance and transcribes their speech into text. Transcription of speech makes pronunciation errors visible, which allows learners to easily notice their errors. The app also scores learners' speech. Those visible feedbacks activate learners' noticing. In other words, ASR help learners better notice the gap between their pronunciation and target pronunciation, so they can spare more time to practice problematic items and correct them step by step. Training, noticing and correcting form a positive circle, which improves learners' pronunciation spirally. The finding follows the idea of 'noticing' highlighted by Schmidt (1990). For low-proficient learners, they have poor pronunciation awareness and more pronunciation errors. They cannot easily find out their pronunciation errors by themselves, but the explicit feedback given by ASR can help them notice those errors. Then learners can correct them by imitating and practicing accordingly in the ASR-based APP. However, ASR cannot tell the difference between more subtle pronunciation sounds, so the feedbacks given by ASR cannot activate enough noticing for high-proficient learners. Furthermore, high-proficient learners equip with better pronunciation at the beginning, so their developmental zone may be limited, which causes no significant improvement of their pronunciation.

5 Conclusion and Implications

The present study conducted a 28-day reading aloud experiment with Chinese senior high students by using ASR-based app *Oral English Drill & Test* to testify the specific effect of ASR on pronunciation accuracy. The results indicated that ASR-supported reading aloud practices have a positive influence on improving pronunciation accuracy and learners accept this ASR-aided training. The effect of ASR-supported training on pronunciation improvement varies with learners' language proficiency. It helps low-proficient learners improve overall pronunciation accuracy and correct word error and deletion error, while it cannot prove to be effective in improving overall pronunciation accuracy of high-proficient learners but can help them correct deletion error and addition error. Hence, Chinese EFL learners can use the app to practice pronunciation autonomously and independently in daily learning to correct some common errors. As Fouz-González (2015) proposes that pronunciation training must be tailored to the learners' needs and technology should be understood as a facilitator to address the most problematic aspects, when using ASR to facilitate pronunciation training, we should take learners' proficiency into consideration. Low-proficient learners can make full use of the ASR tool to improve their pronunciation by correcting their serious pronunciation errors. High-proficient learners can improve their pronunciation habits like deleting and adding sounds involuntarily. It should be noted that ASR still cannot correct all pronunciation errors and cannot tell the difference between some subtle phonemes like [v] and [w], [r] and [l]. Thus, teachers should pay attention to those subtle phonemic errors and give explicit instruction on how to pronounce them. The combination of form-focused training and explicit instruction can foster more positive gains in pronunciation (Couper 2011; Saito 2013). In ASR-supported training, pronunciation may achieve greater improvement if added some explicit instruction.

Acknowledgements. This work is supported by the Center for Language Cognition and Assessment, South China Normal University. It's also the result of the following funds: Guangdong Academic Degrees & Graduate Education Reform Project (2019JGXM37), and University-Industry Collaborative Education Program (201801191006 & 201801191027).

References

Couper, G.: What makes pronunciation teaching work? Testing for the effect of two variables: socially constructed metalanguage and critical listening. Lang. Aware. **20**(3), 159–182 (2011)

Derwing, T., Munro, M.: Second language accent and pronunciation teaching: a research-based approach. TESOL Q. **39**, 379–397 (2005)
Ehsani, F., Knodt, E.: Speech technology in computer-aided language learning: Strengths and limitations of a new CALL paradigm. Lang. Learn. Technol. **2**(1), 54–73 (1998)
Elimat, A.K., AbuSeileek, A.F.: Automatic speech recognition technology as an effective means for teaching pronunciation. JALT CALL J. **10**(1), 21–47 (2014)
Eskenazi, M.: Using automatic speech processing for foreign language pronunciation tutoring: some issues and a prototype. Lang. Learn. Technol. **2**(2), 62–76 (1999a)
Eskenazi, M.: Using a computer in foreign language pronunciation training: what advantages? Calico J., 447–469 (1999b)
Fouz-González, J.: Trends and directions in computer assisted pronunciation training. In: Mompean, J., Fouz-González, J. (eds.) Investigating English Pronunciation: Trends and Directions, pp. 314–342. Palgrave Macmillan, Basingstoke (2015)
Hincks, R.: Speech technologies for pronunciation feedback and evaluation. ReCALL **15**(1), 3–20 (2003)
Howlader, M.R.: A History of English Language Teaching. Oxford University Press, New York (1984)
Hsu, L.: An empirical examination of EFL learners' perceptual learning styles and acceptance of ASR-based computer-assisted pronunciation training. Comput. Assist. Lang. Learn. **29**(5), 881–900 (2015)
Isaacs, T., Trofimovich, P.: Deconstructing comprehensibility: identifying the linguistic influences on listeners L2 comprehensibility ratings. Stud. Second. Lang. Acquis. **34**(3), 475–505 (2012)
Kruk, M.: Using online resources in the development of learner autonomy and English pronunciation: the case of individual learners. J. Second. Lang. Teach. Res. **1**(2), 113–142 (2012)
Levis, J., Suvorov, R.: Automated speech recognition. In: Chapelle, C. (eds.) The Encyclopedia of Applied Linguistics, pp. 316–323 (2014)
Liakin, D., Cardoso, W., Liakina, N.: Learning L2 pronunciation with a mobile speech recognizer: French/y/. CALICO J. **32**(1), 1–25 (2014)
Liakin, D., Cardoso, W., Liakina, N.: Mobilizing instruction in a second language context: Learners' perceptions of two speech technologies. Languages **2**(3), 11 (2017)
Liu, X., et al.: Improving English pronunciation via automatic speech recognition technology. Int. J. Innov. Learn. **25**(2), 126–140 (2019)
Liu, X., Zhu, C., Jiao, J., Xu, M.: Promoting English pronunciation via mobile devices-based automatic speech evaluation (ASE) technology. In: Cheung, S.K.S., Kwok, L., Kubota, K., Lee, L.-K., Tokito, J. (eds.) ICBL 2018. LNCS, vol. 10949, pp. 333–343. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-94505-7_27
Luo, L., Zhang, L.: Review of the history of English phonetic teaching and suggestions on present English phonetic teaching. Foreign Lang. Their Teach. **10**, 21–23 (2002)
McCrocklin, S.M.: Pronunciation learner autonomy: the potential of automatic speech recognition. System **57**, 25–42 (2016)
Morley, J.: The pronunciation component in teaching English to speakers of other languages. TESOL Q. **25**(3), 481–520 (1991)
Neri, A., Cucchiarini, C., Strik, H.: Effective feedback on L2 pronunciation in ASR-based CALL. In: Proceedings of the Workshop on Computer Assisted Language Learning, Artificial Intelligence in Education Conference, pp. 40–48 (2001)
Neri, A., Cucchiarini, C., Strik, H.: Automatic speech recognition for second language learning: how and why it actually works. In: Proceedings of ICPhS, pp. 1157–1160 (2003)
Neri, A., Cucchiarini, C., Strik, H.: The effectiveness of computer-based speech corrective feedback for improving segmental quality in L2 Dutch. ReCALL **20**(2), 225–243 (2008)

Pennington, M.C.: Computer-aided pronunciation pedagogy: promise, limitations, directions. Comput. Assist. Lang. Learn. **12**(5), 427–440 (1999)

Saito, K.: Reexamining effects of form-focused instruction on L2 pronunciation development: the role of explicit phonetic information. Stud. Second Lang. Acquis. **35**(1), 1–29 (2013)

Schmidt, R.W.: The role of consciousness in second language learning. Appl. Linguist. **11**(2), 129–158 (1990)

Sharifi, M., Azizifar, A., Jamalinesari, A., Gowhary, H.: The effect of rosetta stone computer software on vocabulary learning of Iranian elementary EFL learners. Procedia Soc. Behav. Sci. **192**, 260–266 (2015)

Tsai, P.H.: A duet of pedagogy and technology—an evaluation of my ET, a computer assisted pronunciation training system made in Taiwan. In: Proceedings of the 21st International Conference on English Teaching and Learning in the ROC, pp. 439–452 (2003)

Tsai, P.H.: Beyond self-directed computer-assisted pronunciation learning: a qualitative investigation of a collaborative approach. Comput. Assist. Lang. Learn., 1–28 (2019)

Wang, Y.: Reflection on teaching English phonetics in middle schools from the perspective of university freshmen's phonetic competence. Curric. Teach. Mater. Method **30**(07), 56–60 (2010)

Xu, Y., Zeng, Y.: How to deal with students' pronunciation errors—a study based on the corpus of test-takers' performance in computer-based English listening and speaking test of Gaokao (Guangdong version). Curric. Teach. Mater. Method **35**(09), 90–96 (2015)

Improving English Phoneme Pronunciation with Automatic Speech Recognition Using Voice Chatbot

Victor Chi-Wa Cheng[1(✉)], Vincent King-Tang Lau[2], Ringo Wai-Kit Lam[3], Tian-Jie Zhan[1], and Pak-Kin Chan[3]

[1] Zhi Yu Technology (Jiangmen) Ltd., Jiangmen, China
victor.gmbox@gmail.com
[2] Doctors Think Tank Academy, Kowloon, Hong Kong
vincent.lau@dttac.com
[3] AITalk IT Ltd., Kowloon, Hong Kong
ringo@aitalkit.com

Abstract. Computer Aided Pronunciation Training (CAPT) is gaining attention in recent years especially for Second Language (L2) and Foreign Language (FL) learners. Nevertheless, operating such learning tools usually requires some training which may hinder the users' interest. To address these issues, we propose a Phonics Learning Voice Chatbot (PLVC) using Automatic Speech Recognition (ASR) and Triplet Neural Networks (TNNs). TNNs have excellent capabilities in discerning different entities that are similar unless examined in fine grain scales. This distinctive feature enables them to be utilized in classifying and assessing phonemes. The employment of chatbot in our teaching process not only relieves users from using keyboard, mouse, and menus, the teaching storyline can also be adjusted automatically and imperceptibly based on the user replies during teaching. Finally, PLVC has one unique feature that it can be used to assess the performance of the users in phonic learning. We performed an experiment with 10 users and compared the ratings given by PLVC and experienced phoneme assessors and found that the correlation coefficient is 0.71, showing the assessment done by PLVC is reliable.

Keywords: Computer Aided Pronunciation Training (CAPT) · Voice Chatbot · Speech recognition · Triplet neural network

1 Introduction

Mastering phonemes of a second language (L2) or a foreign language (FL) is very important for learners [1, 10, 14], as poor pronunciation may inhibit understanding by listeners. Although not better than human instruction, Computer Assisted Pronunciation Training is an excellent medium for attaining near-native pronunciation for most users. CAPT can be used in two main areas: (1) pronunciation assessment, and (2) pronunciation learning/teaching, where the users are usually students and teachers.

L.-K. Lee et al. (Eds.): ICTE 2020, CCIS 1302, pp. 88–99, 2020.
https://doi.org/10.1007/978-981-33-4594-2_8

Recently, due to the epidemic of Covid-19, most of the normal classroom teaching activities including English phonics learning are stopped and replaced by distance learning such as video conferencing or video watching. However, distance learning is believed not as effective as the normal classroom teaching. Teachers cannot easily take care of all students' responses with a compact screen. Besides, students may sometimes respond simultaneously and cause a mess and they are also easily distracted by family members. One-on-one teaching can reduce the problem but limited teaching time and high cost hinders this learning mode. In this regard, CAPT is obviously an excellent alternative.

1.1 Background

There are papers summarizing the research and develop efforts in CAPT [5, 15]. Recently, Agarwal et al. [1] reviewed the tools and techniques for CAPT in English. Research [6] showed that pronunciation training to imitate a chosen well-matched native speaker was positive to L2 and FL learners, reducing foreign accentedness.

Enunciate [14] is an online CAPT system for Chinese learners of English. It consists of an audio-enabled web interface, a speech recognizer for mispronunciation detection and diagnosis, a speech synthesizer and a viseme animator. Based on it, Qian et al. [13] proposed a two-pass framework with discriminative acoustic modeling for mispronunciation detection and diagnoses (MD&D). As reported in the literature, when conducting detection and diagnosis consecutively in the two-pass framework, discriminative training of two sets of HMMs can reduce the Phone Error Rate down to 16.5% where it is calculated by first aligning the manual transcription with the recognition transcription and extracting the mismatches. On the other hand, Ai-Barhamtoshy et al. [3] have proposed Speak Correct system which has four main stages, trainer, decoder, evaluation and phonetic editor. Ai and Xu [2] also introduced a CAPT framework for L2 learners with the capability of supporting iterative improvement of automatic pronunciation error recognition and classification.

Goodness of pronunciation (GOP) is the most widely used method for mispronunciation detection in CAPT, Huang [9] proposed a transfer learning approach to GOP based mispronunciation detection. Jonas [10] explored the English File Pronunciation (EFP) app to help FL learners improve pronunciation.

1.2 Phonics Learning Voice Chatbot

In our study, we use Automatic Speech Recognition (ASR) and Triplet Neural Networks (TNNs) [8] to build the CAPT and our system is named as Phonics Learning Voice Chatbot. TNNs were first proposed by Google and have shown to be efficient for classification tasks with only a few training samples per class and later Cheuk [4] used TNNs in speech on Speaker Detection and Identification. Our study also showed that TNNs can have good results in phoneme classification.

2 Deep Neural Networks

Most existing classification approaches usually require "good" features derived from the training samples and use a classifier to classify test samples. Nevertheless, those features may not be easily identified and sometimes it may require domain experts or even heuristics. On the other hand, deep neural networks [12], neural networks with many layers (usually more than 4 layers), have a desirable property that they are able to jointly learn features and classifiers simultaneously given the samples, while only the structures of the networks needed to be determined by humans.

Deep learning models have achieved great success on many tasks that are previously having modest performance. For example, image classification and speech recognition are now very applicable and ubiquitous. Phoneme classification and assessment, however, are different tasks because some phonemes are very close to each other and easily be mixed or mis-classified, e.g., /r/ and /l/, and fine-grained features are required to correctly discern them. It may be prohibitive to manually derive such features and we propose the use of the triplet neural networks as they address both the similarity between same category phonemes and the dissimilarity between phonemes coming from different categories.

2.1 Triplet Neural Networks (TNNs)

Briefly, TNNs [8] have 3 identical neural networks (identical in structure and weights but with different names) known as positive network, anchor network, and negative network. The structure of the neural networks depends on applications. For our tasks of phoneme classification and assessment, we use convolution layers and full connection layers to build the networks (described later). During a training instance, 3 distinct samples are selected from the training data with the requirement that the sample for the positive network input should be more similar to the sample presented to the anchor network than the sample presented to the negative network. With this setting, a TNN should map the 3 input samples (after training) to a feature space such that positive sample is closer to anchor sample than negative sample, under a metric (distance) function, e.g., L2 norm. Figure 1 shows a typical TNN.

In Fig. 1, $\boldsymbol{p}$, $\boldsymbol{a}$, and $\boldsymbol{n}$ are vector outputs of the neural networks with respect to the positive input, anchor input, and negative input. The loss of the TNN is computed with the triplet cost function given by (1):

$$Loss = max(d(p, a) - d(n, a) + margin, 0.0), \tag{1}$$

where $d(\cdot)$ is a distance function and *margin* is a positive number. It is not difficult to see that the loss function is minimum when the feature distance between the anchor input and the positive input is 0.0 (i.e. $d(p, a) = 0$) while the feature distance between the anchor input and negative input is greater than margin (i.e. $d(n, a) > margin$). When compared the triplet loss function with the cross-entropy function used in conventional classification neural networks, the former enables TNNs to learn a feature mapping that pay more attention on fine-grained details between the inputs, especially the similarity between the anchor and positive inputs and the dissimilarity between anchor and negative inputs. As a result, TNNs are more suitable for classification task where the entities are only different from each other in small details. Although TNNs has this desirable feature, employing triplet loss function in training may cause the process slower and sometimes not convergent.

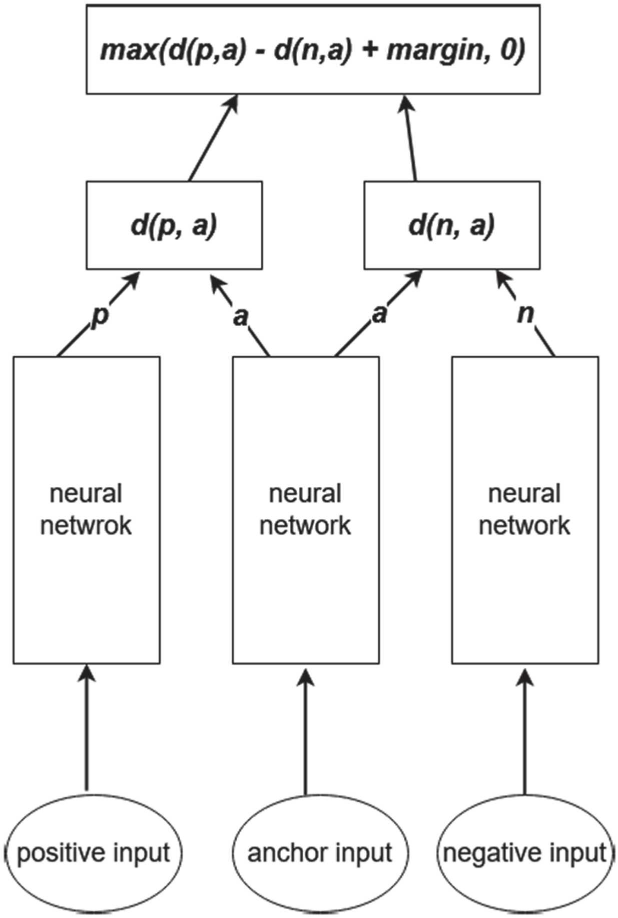


Fig. 1. Structure of a typical TNN

2.2 Network Structure

Figure 2 shows the structure of the neural network in the TNN used in our phoneme classification and assessment.

The input of the network is the Mel scale transformed spectrograms obtained by preprocessing the input audio signals. A spectrogram can be regarded as a grayscale image with discrete frequencies as the vertical axis whereas the discrete time as the horizontal axis. The brightness of pixel at coordinate (f, t) represent the power of the signal at frequency f and time t. Figure 3 shows a spectrogram of the phoneme /b/ pronounced by a person.

The transformed spectrogram goes through two consecutive convolution layers (CNN) with maximum pooling. The pooling outputs then go through two fully connected layers and become 9-dimensional vectors where the loss is computed and then weights of the neural networks are adjusted by using the Adam Optimization Algorithm.

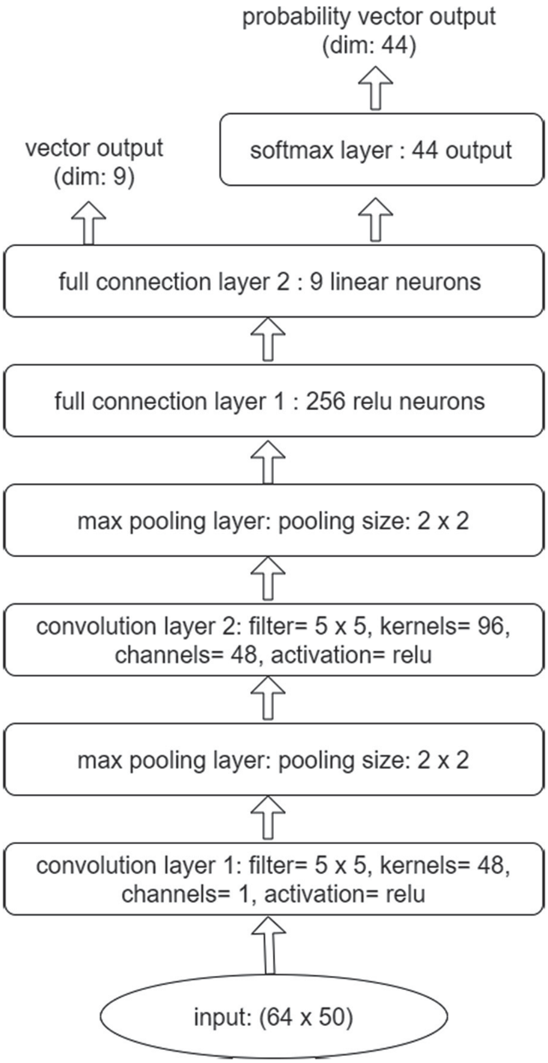


Fig. 2. Structure of the neural network used in the TNN

If training does not converge, the learning rate should be reduced. In our experiments, the learning rate is set to 0.001.

After training, the TNN should be able to map unseen phonemes pronounced by users to the 9-dimensional space such that phonemes having same category should be close to each other and categories of phonemes that having close pronunciations should also be close together while categories of phonemes having distinctive differences should be far apart from each other.

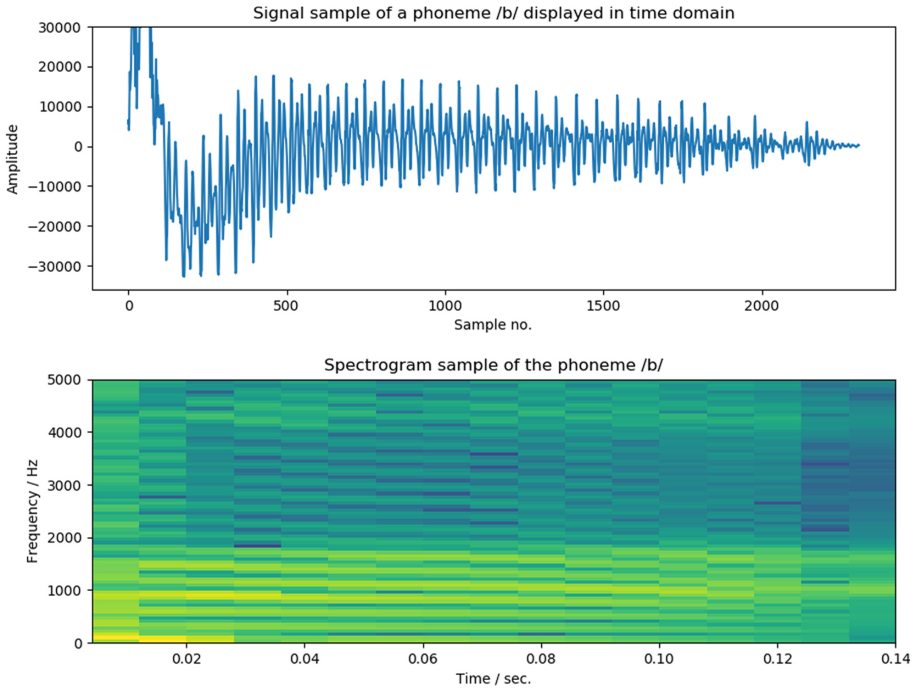


Fig. 3. A sample of a phoneme /b/ displayed in time domain and spectrogram. Note that the frequency axis is in linear scale (not Mel scale) for better visualization

If a set of high-quality phoneme pronunciations is available, they can serve as the references and the neural network output vector distance between the references and a user pronunciation can be used to rate/compute the quality of the pronunciation of users.

For classifying phonemes, an additional softmax neural network layer should be added on top of the neural network output (referring to the right top portion of Fig. 2). After training, it can classify phonemes at the softmax layer which has 44 outputs corresponding to 44 phonemes. The value of each output gives the probability of that phoneme.

2.3 Input Preprocessing

In preprocessing of the phoneme audio signals, we follow classical methods which parameterizes the signals into a set of feature values. Most of the methods generate the features based on analyzing the power distribution in frequency domain during a short time interval, known as the window. The most popular method is the Mel frequency cepstral coefficients (MFCC) [11], in which a signal is first divided into small overlapping windows and each window undergoes a power spectrum analysis subjected to a Mel-based transform of the frequency axis. Then the obtained values of Mel-bands are taken logarithm and follow by decorrelation with discrete cosine transform. This approach is quite standard in automatic speech recognition using conventional GMM-HMM [7]. Nevertheless, it is found that the correlated property of the features is helpful in neural networks with convolution layers. Thus, the discrete cosine transform is usually not performed and the Mel scale frequency spectral coefficients are directly treated as the features. Figure 3 illustrates a spectrogram sample (in linear frequency scale) of phoneme /b/. It is clear that the main power concentrates on low frequencies at the beginning of the pronunciation.

In our tasks, the sampling frequency for input audio is 22.05 kHz. We set the lower and upper frequencies of analysis to be 0 Hz and 8 kHz. The Mel scale of this frequency range is then divided into 64 points. Maximum audio signal length for a single phoneme is 2 s. Audio signals less than 2 s. will be padded with zeros to 2 s. length. The window size for computing spectrograms is 100 ms with step size of 40 ms. Hence, the spectrogram will have size 64×50 and this size match the neural network shown in Fig. 2.

3 Voice Chatbot for Phoneme Classification and Assessment

Having the TNN as the core for phoneme classification and assessment, a chatbot is developed which highly simplify the interaction between the users and the computers. Traditionally, users mainly use keyboard, mouse, menus, etc. to interact with CAPT system and a short training on operating the system is often required. The use of chatbot can minimize such hindrances by using voice as the communication medium. This not only simplifies the task but also strengthens the interest of users. Figure 4 presents the flow of the task using voice chatbot.

The task flow is designed such that users can interact with the system with voice instead of using mouse and keyboard. Furthermore, challenging voice recognition can now be reduced to “yes-no” voice recognition which makes the recognition simple and robust. The output of the chatbot can be strings with pre-recorded voices or animations to make the learning process more interesting. Thus, the system operation is not only easy for adults, but also suitable for children and even for some users with disability. Figure 5 illustrates the system’s main screen.

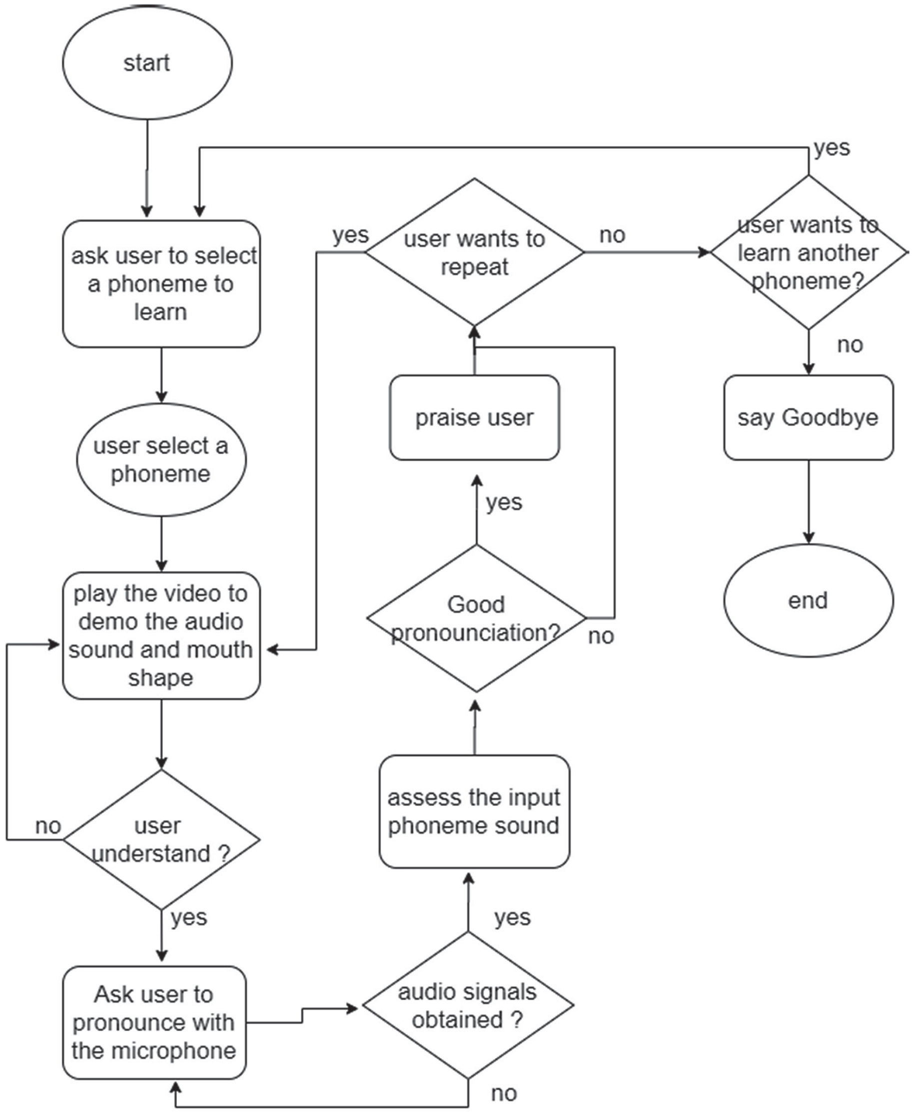


Fig. 4. Flowchart for phoneme classification and assessment. Main screen of the system is illustrated in Fig. 5

4 Experiments and Results

We collected audio samples where each sample represents one phoneme sound of one person. There are 2 data sources. The first source is the Professional source, from which we collected 1862 valid samples from 6 English teachers as our anchor samples for the TNNs. The second source is from a group of 10 amateur users and they recorded their pronunciation of the 44 phonemes. Each person makes at least 10 samples for each

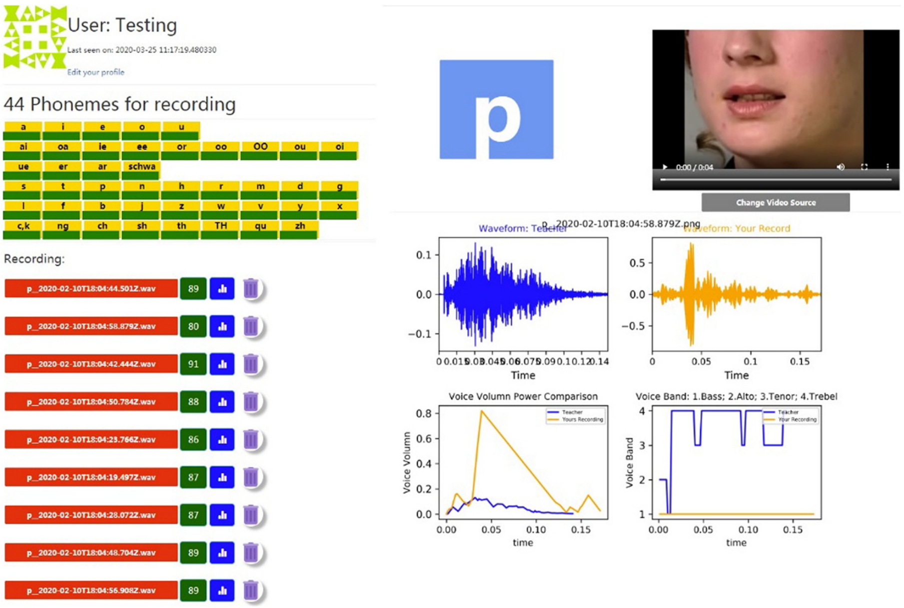


Fig. 5. Main screen of the system for phoneme classification and assessment.

phoneme. Totally, we have 4893 samples, after removing those invalid samples such as noise, corrupted or wrong pronunciation. Table 1 shows the sample summary for this experiment.

Table 1. Number of samples for the experiment

Sources	# of persons	# of valid samples
Professional (anchor)	6	1862
Amateur user (example)	10	4893

4.1 Experiment

In the preprocessing stage, we removed the silence header and ending section for each sample. The samples of professionals will be taken as anchors and the other will be taken as examples. For each anchor and example, we extract 64 Mel Scale Frequency Spectral coefficients as the input features for the TNN training. We grouped some similar phonemes into phoneme clusters where even the human is difficult to discern the phonemes within a cluster, as shown in Table 2. Based on this clustering setting, we categorize the 44 phonemes into 23 unique labels where 5 labels are cluster labels and the rest 18 labels are single phoneme classes. For samples in each cluster, there will be a second classification, using k-nearest neighbors (KNN) classifier, to classify them into individual phoneme classes.

Table 2. Phonemes clusters

Cluster labels	Phoneme
ngnmvthzzh	/n/,/m/,/ng/,/v/,/th/,/z/,/zh/
gdbcktpwqsw	/g/,/d/,/b/,/c, k/,/t/,/p/,/qu/,/schwa/,/w/
sshfth2	/s/,/sh/,/f/,/TH/
rerlor	/r/,/er/,/l/,/or/
uar	/u/,/ar/

Thus, the input are samples of vectors with 64 entries and the output are 9-dimensional vectors and probabilities of 23 labels. We combined 1862 anchor samples and 4000 amateur samples to form the training set and the rest 893 amateur samples are treated as the test set. After training, we remapped the output of each sample into 2-dimensional space by using t-SNE visualization tool to illustrate the separability of the 23 labels of phonemes.

4.2 Results

Figure 6 illustrates the classification results. For phoneme clusters, we ran a second phrase classification. Figure 7 shows the classification results on 3 phoneme clusters. The overall classification accuracy is 85% when evaluated with the test dataset.

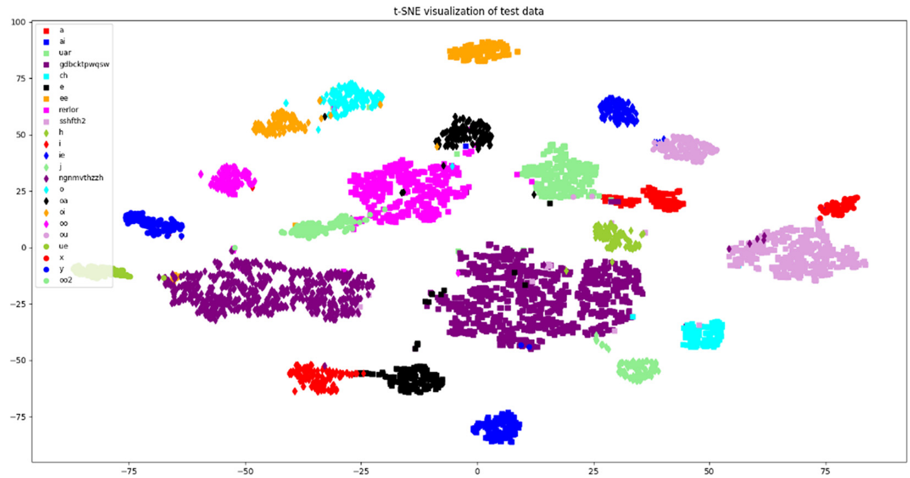
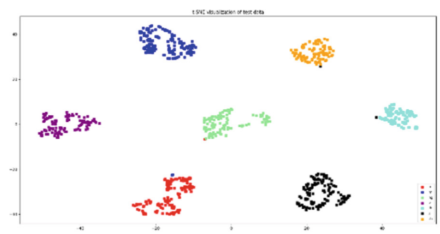
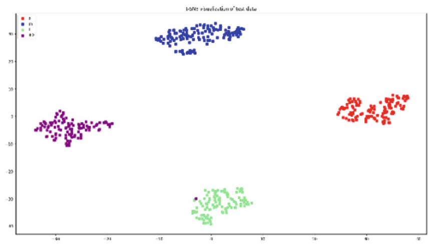


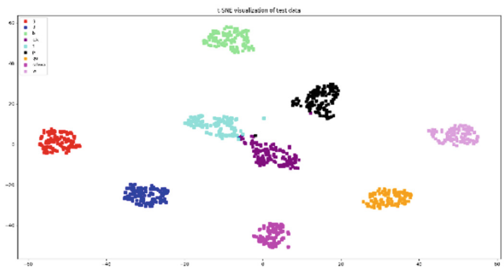
Fig. 6. 2-D overview of the distribution of the 23 phoneme labels



(a) T-SNE visualization of phoneme /n/,/m/,/ng/,/v/,/th/,/z/,/zh/



(b) T-SNE visualization of phoneme /s/,/sh/,/f/, /TH/



(c) T-SNE visualization of phoneme /g/, /d/, /b/, /c,k/, /t/, /p/, /qu/, /schwa/, /w/

Fig. 7. T-SNS visualization of classification results of three phoneme clusters

Given a phoneme class, we can assess the quality of a pronunciation, uttered by a user, by computing the distance between its TNN output and the mean of the anchor outputs, say *d1*. We also pre-compute the distances between the training sample outputs and the mean of the anchor outputs and obtain the median distance for that phoneme, say *dm*. With |*d1* – *dm*|, we can score the user pronunciation based on a 3-levels scoring scheme: Good, Average and Bad.

For testing the assessment performance of our system, we invited 3 professional English teachers to manually score the 893 test samples into Good, Average and Bad labels. In comparing the results generated by our system and the human benchmark, we found that coefficient of correlation between them is 0.71. This is a strong support that our system is effective in classifying and assessing phoneme pronunciations.

5 Conclusions

With the use of Automatic Speech Recognition and Triple Neural Networks, we propose a Phonics Learning Voice Chatbot which enables phoneme pronunciation learning through speech instead of using keyboard, mouse or touch panel. We compared its scoring results with professional English teachers' scores based on a 3-levels scheme: Good, Average and Bad. It is found that the scores given by the system are highly correlated with the scores from the professional English teachers.

References

1. Agarwal, C., Chakraborty, P.: A review of tools and techniques for computer aided pronunciation training (CAPT) in English. Educ. Inf. Technol. **24**(6), 3731–3743 (2019). https://doi.org/10.1007/s10639-019-09955-7
2. Ai, R., Xu, F.: A system demonstration of a framework for computer assisted pronunciation training. In: Proceedings of ACL-IJCNLP (2015)
3. Ai-Barhamtoshy, H., Jambi, K., Ai-Jedaibi, W., Motaweh, D., Abdou, S., Rashwan, M.: Speak correct: phonetic editor approach. Life Sci. J. **11**(8), 626–640 (2014)
4. Cheuk, K., Balamurali, B.T., Roig, G., Herremans, D.: Latent space representation for multi-target speaker detection and identification with a sparse dataset using triplet neural networks. In: Proceedings of the IEEE Automatic Speech Recognition and Understanding Workshop (2019)
5. Eskenazi, M.: An overview of spoken language technology for education. Speech Commun. **51**(10), 832–844 (2009)
6. Felps, D., Bortfeld, H., Gutierrez-Qsuna, R.: Foreign accent conversion in computer assisted pronunciation training. Speech Commun. **51**(10), 920–932 (2009)
7. Gales, M., Young, S.: The application of hidden Markov models in speech recognition. Found. Trends Signal Process. **1**(3), 195–304 (2007)
8. Hoffer, E., Ailon, N.: Deep metric learning using triplet network. In: Feragen, A., Pelillo, M., Loog, M. (eds.) SIMBAD 2015. LNCS, vol. 9370, pp. 84–92. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-24261-3_7
9. Huang, H., Xu, H., Hu, Y., Zhou, G.: A transfer learning approach to goodness of pronunciation automatic mispronunciation detection. Acoust. Soc. Am. **142**(5), 3165–3177 (2017)
10. Jonas, F.G.: Using apps for pronunciation training: an empirical evaluation of the English file pronunciation app. Lang. Learn. Technol. **24**(1), 62–85 (2020)
11. Muda, L., Begam, M. and Elamvazuthi, I.: Voice recognition algorithm using Mel frequency cepstral coefficient (MFCC) and dynamic time warping (DTW) Techniques. J. Comput. **2**(3). arXiv:1003.4083 (2010)
12. Nielsen, M.A.: Neural Networks and Deep Learning. Determination Press (2018)
13. Qian, X., Meng, H., Soong, F.: A two-pass framework of mispronunciation detection and diagnosis for computer-aided pronunciation training. IEEE/ACM Trans. Audio Speech Lang. Process. **24**(6), 1020–1028 (2016)
14. Yuen, K., Leung, W., Liu, P., Wong, K., Qian, X., Lo, W., Meng, H.: Enunciate: an internet-accessible computer-aided pronunciation training system and related user evaluations. In: Proceedings of the 2011 International Conference on Speech Database and Assessments (Oriental COCOSDA), Hsinchu, Taiwan (2011)
15. Witt, S.M.: Automatic error detection in pronunciation training: where we are and where we need to go. In: International Symposium on Automatic Detection on Errors in Pronunciation Training (ISADEPT) (2012)

The Design and Implementation of a Computerized Adaptive Testing System for School Mathematics Based on Item Response Theory

Jiyou Jia(✉) and Huixiao Le

Department of Educational Technology, Graduate School of Education, Peking University, Yiheyuanlu 5, Beijing 100871, China
{jjy,lehuixiao}@pku.edu.cn

Abstract. On the contrary to classical school test, the computerized adaptive test (CAT) based on IRT (Item Response Theory) stepwise estimates every examinee' response to every question item and provides the examinee with the next corresponding and appropriate question item. The online intelligent tutoring system "Lexue 100" has been used by more than 100,000 junior high school students in China in the past 7 years. It has more than 70,000 sets of mathematics quizzes designed for different units in the various textbooks. Each of about 68,600 quizzes among the 70,000 sets has been used by more than 100 students. We analyzed those quizzes, calibrated the original 740,910 IRT parameters for the questions composing the quizzes, recalculated the parameters by using linking algorithms for 194,205 question items. Based on this large question bank with pre-calibrated parameters for every question, we designed and implemented a computerized adaptive test system for school mathematics.

Keywords: Computerized adaptive test · Intelligent tutoring system · School mathematics · Item response theory · Classical test theory

1 Introduction

1.1 The Classical Test Theory (CTT)

The classical test theory (CTT) believes that an examinee's observed or obtained score in a test is the sum of a true score or error-free score and an error score (Novick 1966; Lord and Novick 1968; Allen and Yen 2002). It refers to the traditional theory and methods that is easy to apply in practical testing situations in schools and higher education institutes (Hambleton and Jones 1993). Therefore it has served the measurement community for a long time.

Focusing on test-level information such as the reliability exemplified by Cronbach's Alpha under rather mild assumptions and the validity of a test, CTT model also studies the statistics of the components in the test, i.e., the test items, including item difficulty and

L.-K. Lee et al. (Eds.): ICTE 2020, CCIS 1302, pp. 100–111, 2020.
https://doi.org/10.1007/978-981-33-4594-2_9

item discrimination. CTT does not use a complex theoretical model to link examinee's ability with the success rate of a special item. On the contrary, CTT considers a group of examinees and empirically examines their success rate on the item by assuming that the item is dichotomous scored.

The success rate of a certain examinee on a certain item is called the p (proportion) value of the item, which is used as an indicator of the difficulty of the item. But in fact, it is a reverse indicator of item difficulty, and the higher the value, the easier the item is. The difficulty of an item is calculated as the following formula, and has a value in the set [0, 1].

$$p = \frac{\text{number of examinees whose answer are correct}}{\text{number of all examinees}}$$

The ability of an item to distinguish subjects with higher ability and those with lower ability is called item discrimination, and is usually expressed statistically by Pearson product moment correlation coefficient between the scores on the item, for example, 1 and 0 of the right and wrong marks on an item respectively, and the total test scores. If an item is scored by dichotomy, the discrimination is usually calculated by the correlation coefficient of point dichotomy, i.e. the point-biserial correlation, r_{pb}. It is equal to the Pearson product moment correlation r_{XY}, if there is a continuously measured variable X, i.e. the test score, and a dichotomous variable Y, i.e. the dichotomous score in one test item, 1 for right or 0 for false, as shown in the following formula.

$$r_{pb} = \frac{\mathrm{M1} - \mathrm{M0}}{\mathrm{Sn}} * \sqrt{\frac{\mathrm{N1N0}}{N^2}}$$

where $M_{1,}$ M_0 is the average value on the continuous variable X for all data points in group 1 with the test score 1 and group 2 with the test score 0, respectively, $N_{1,}$ N0 is the number of data points in group 1 and 2 respectively, Sn is the standard deviation of the total sample, and N is the total sample size.

In this sense, the classical test theory can be regarded as almost synonymous with the true score theory. The main limitation of CTT can be summarized as the cyclic dependence: the human statistics, i.e. observed scores, are item dependent, while item statistics, including item difficulty and discrimination, depend on candidates. This cyclical dependence brings some theoretical difficulties to the application of CTT in some measurement situations. In other words, the ability of examinees to participate in a test is determined by the difficulty of the test, and the difficulty of the test is determined by the ability of the examinee taking the test.

1.2 The Item Response Theory (IRT)

The term "classical" refers not only to the birth age of these models, but also contrasts to the relatively new psychometric theory, which is generally referred to as item response theory (IRT), which is sometimes described as modern latent theory. Item response theory mainly focuses on item level information, in contrast to CTT focusing on test level information. It describes dichotomous items as subjects' response variables to questionnaires or test items. These responses are modeled as a function of a latent trait, such as the ability level. According to the theory, some item parameters related to project characteristics can be calculated, such as difficulty, discrimination and guessing. The IRT models are used to measure the investigated latent traits (van der linden and Hamilton 1997; de Ayala 2009).

According to IRT, the item and ability parameters remain unchanged under linear transformation. The mean and variance estimates of different subgroups can be changed so that they are on the same scale. Except for sampling or measurement errors, the estimated values of subjects' ability will be independent of the given item pool. One advantage of IRT models is that the ability and item parameters remain invariant. Because IRT model provides the information at the item level, IRT model is much better than CTT model in many test applications, especially in those specially designed to investigate the performance of the candidates on single test items.

IRT has two main hypotheses. The first is that the examinees' performance in test items is a function of potential traits or abilities. The second is that the relationship between examinees' potential traits and their probability of answering items correctly exists in the form of a monotonic increasing function, whose graph is called item characteristic curve (ICC). With the growth of latent traits, the probability of answering items correctly will either increase or remain unchanged (Hambleton et al. 1991). In IRT, an underlying latent trait is referred to as theta (θ), which is conceptually similar to a true score in CTT.

The graph of an item characteristic curve (ICC) has the latent trait theta (θ) on its x-axis, typically expressed as a Z-score with the range from -3 to $+3$, and the probability of a correct response (PCR) on its y-axis ranging from 0 to 1.

The IRT framework is composed of a set of models. The application of each model in a specific situation depends on the nature of the test items and the feasibility of different theoretical assumptions about the test item. There are three IRT models for the dichotomously scored test items, namely, one parameter, two parameter and three parameter IRT models.

The one parameter logic model (1PL) uses the difficulty of the item, i.e. "b" parameter to explain the relationship between the ability level and the correct answer probability of the item. It is the point on the ability rating corresponding to the position with a probability of correct answer of 0.5 on ICC.

Two parameter logic model (2PL) not only uses the b parameter, i.e. item difficulty, but also adds another parameter, which indicates how a project divides candidates into different ability levels. It is called "a" parameter or item discrimination parameter, which is equal to the slope of ICC apex.

The three-parameter logic model (3PL) is based on the two parameter model, adding the pseudo chance level "c" parameter or guessing parameter, which represents the probability of candidates with low ability score to correctly answer an item. It is the value of the lower asymptote of the ICC.

No matter which IRT model is used, the calculation and determination of the parameters are the most important procedures, which are called the parameter calibration.

Although these three item parameters are useful in themselves, they are combined into an item information function (IIF) for the purpose of computerized adaptive testing (CAT). It describes the measurement effect or accuracy of an item at each level of the trait measured by a particular test. One of the main advantages of IRT is that both the item and the examinee are put on the same scale (usually a standard score scale with mean = 0.0 and standard deviation = 1.0), so that people can be compared with items and vice versa.

The IIF is calculated from item parameters. The IRT usually uses maximum likelihood estimation (MLE) to estimate candidates' θ level and parameter calibration. The maximum likelihood estimation method weights each item by item parameters and considers whether the examinee answers each item correctly. Due to the combination of the whole answer pattern and the characteristics of each item, MLE can provide more detailed differentiation of candidates than CTT's number-correct scoring. For example, if a traditional test with three items is scored correctly, up to four scores (0 to 3) can be obtained, while MLE can obtain eight (2^3) different theta estimates for the same test. MLE also provides a personalized standard error of measurement (SEM). The SEM can vary from person to person, depending on how they answer a specific set of items.

In 3PL, the probability of a correct response to item i for examinee j, p, is a function of the three item parameters and examinee j's true ability $\theta_{j,}$ as shown in the following formula.

$$\mathrm{P}\left(u_i = 1 \mid \theta_j,\ a_i,\ b_i,\ c_i\right) = c_i + \frac{1 - ci}{1 + \mathrm{e}^{-1.7\mathrm{ai}(\theta\mathrm{j} - bi)}}$$

1.3 Parameter Equating and Linking

In most test scenarios, for safety reasons, test forms are required to be composed of different items, so that test scores are not comparable in different tests. Traditionally, the term equating refers to adjusting the scores of parallel forms in order to make them as similar in content and statistical characteristics as possible (Lee and Ban 2009; Kolen and Brennan 2014). The equating process allows the comparison of scores obtained from different tests. Various statistical methods have been proposed for equating, such as horizontal and vertical scaling.

If the parameters of IRT model are estimated for each group of candidates taking different test forms, they are not comparable due to different sources of measurement scales. But if two tests represent a subset of a common item, the parameters can be set to the same scale. This is equating and linking according to common items. There are two ways to do it. Concurrent or simultaneous calibration includes simultaneous estimation of item parameters and direct measurement results on a common metric. The other method is to calibrate and estimate the item parameters of each test form separately. The item parameter estimation of common item can be used to estimate scale conversion. Both methods can be extended to the case of multiple test forms, but only if these test forms can be linked through a common item. The advantage of simultaneous calibration is that the estimated parameters do not need to be converted because they have been expressed on the same scale. However, this method needs to combine the data of each form into a single data set. If there are many tables or each table is tested on a large number of examinees, the data set becomes very large. In some cases, the large dimensions of the dataset can make estimation very slow or even infeasible. On the contrary, by equating the coefficients of each calibration, the previously managed parameters do not need to be estimated, thus a single dataset does not need to be constructed.

The transformation of parameter estimation is realized through linear transformation, and the transformation coefficients are called equating coefficients. For each pair of tests with common terms, the direct equating coefficients can be calculated. Generally speaking, if two tests can be connected by a series of form chains presenting the common terms in pairs, indirect equating coefficients can be calculated, and the parameter estimation of one test can be transformed into the scale of another term by using linear transformation. These coefficients are functions of direct equating coefficients. Moreover, some linking designs are quite complicated. The two tests may be connected directly or by different chains. For each path connecting the same two tests, the equivalent coefficient can be calculated to produce different scale conversion. If so, the equivalent coefficients can be averaged to obtain a single transformation.

For the calculation of direct equating coefficients, four implementation methods based on item parameters are usually used: mean sigma (Marco 1977), mean (Loyd and Hoover 1980), and response function methods, including Haebara (1980) and Stocking Lord (1983). The bisector method was used to average the coefficients obtained by different paths.

1.4 The Computerized Adaptive Testing (CAT) Based on IRT

Computerized adaptive testing (CAT) is the design of psychological and educational measuring instruments for interactive delivery between one computer and an examiner. CAT can be used for tests of examinees' ability or achievement and for assessments of personality and attitudinal variables. The CAT is hoped to dynamically select the set of test questions from a pre-calibrated item bank for each examinee that can effectively and efficiently measure that examinee on the trait.

The IRT supports CAT with its pre-calibrated item bank. A given trait needs a relatively large item bank whose items together provide information on all areas of the trait or theta. When the subjects start CAT, they get the same initial estimation of theta. Of course this estimation can be made based on any previous information of the subjects, such as the scores of other tests, grades, and information provided by teachers. An item from the item bank which has not been used in the test and corresponds to the initial theta estimation is given to the examinee and can be scored by the computer as soon as the examinee submits the answers. As the test continues, the maximum likelihood estimation (MLE) is used to obtain a new θ estimate based on the examinee's answers to all items in the test. The new estimation of θ is used to select the next item, which has not been used in the item bank and provides the most information in the current theta estimation.

A CAT can be terminated with different termination criteria according to different test objectives. In many educational settings, tests are used to select or classify, for example, to classify a person as having knowledge in certain fields, or to select those who will be admitted or employed by a school or college. The purpose is to classify as accurately as possible. To achieve this goal, theta estimation and related SEM are used in the CAT. If a person's estimated θ is higher than the cut-off value, he or she can be classified as higher than the cut-off value on the θ scale. If this condition is met, the test can be terminated. The result of this test will be a set of classifications for a group of subjects with a maximum error rate of 5%, which can be controlled by changing the size of SEM confidence interval around theta estimation.

Sometimes the CAT can be used for counseling purposes. The purpose of this kind of tests is to measure everyone as much as possible. It is best to measure each subject to the desired level of accuracy according to a predetermined SEM level. If the candidate reaches this level, the test is over. This will result in all candidates having equally accurate scores - perhaps defining a new concept of test fairness. If there are enough test items in the item bank distributed properly on the θ scale, and the test is allowed to last long enough, this goal ca be achieved.

In computer programming, the CAT algorithm is usually an iterative process with the following steps.

(1) Given the currently estimated ability level θ, all the items that have not yet been administered are evaluated to determine the best next item, which provides the most information. With IRT, the maximum information can be quantified as the standardized slope of Pi(θ) at θ, as shown in the following formula.

$$I_i(\theta) = \frac{P_i'(\theta)}{P_i(\theta)(1 - P_i(\theta))}$$

where $I_i(\theta)$ is the information function for item I, $P_i(\theta)$ is the probability of a correct response to item i, and $P'i(\theta)$ is the first derivative of $P_i(\theta)$.

(2) The best next item is given to the examinee and the examinee writes the answers.
(3) A new ability is estimated based on the responses to all the administered items.

A modification of the Newton-Raphson iterative method for solving equations is used, sketched in Loyd and Hoover (1980, p. 181). The test begins with an initial estimate, calculates the probability of a correct response for each item, and then adjust the ability estimation to obtain improved agreement of the probabilities and the observed response vector. The process is repeated until the adjustment is extremely small, as the following formula shows, where the right hand side of the equation is the adjustment, θ_{S+1} denotes the adjusted ability estimate. The denominator of the adjustment, i.e., the sum of the item information functions evaluated at θ_S, is the test information function, Irr.

$$\hat{\theta}_{s+1} = \hat{\theta}_s + \frac{\sum S\left(\hat{\theta}_s\right)_i}{\sum I_i\left(\hat{\theta}_s\right)}$$

The test information function Irr is inserted in the formula below to obtain the standard error SE(θ):

$$SE\left(\hat{\theta}\right) = \frac{1}{\sqrt{I_{\gamma\gamma}}}$$

(4) Steps 1 through 3 are repeated until an ending criterion is met.

1.5 The IRT and CAT in Practice

According to Traub (1997), the beginning of CTT is associated with Spearman's finding about how to correct a correlation coefficient for attenuation due to measurement error and how to obtain the index of reliability needed in making the correction. Since then CTT has been an influential theory of test scores in the social sciences. However, the theory has been superseded by the more sophisticated models in IRT since 1950s and 1960s. Since 1970s and 1980s, with powerful computers the advantages of IRT has been recognized and become the preferred method for developing scales, especially in so-called high-stakes tests in the United States, including the Graduate Record Examination (GRE) and Graduate Management Admission Test (GMAT) from the Educational Testing Service (ETS). The GMAT is primarily administered as a CAT (Eignor et al. 1993).

Since 1980s, the concepts CTT, IRT and CAT have been introduced in the research community of education, psychology and so on in mainland China. But to our best knowledge after searching the related research databases and examination websites, there has not been any practical CAT based on IRT with the pre-calibrated parameters from previously administered tests in mainland China. The reason may be the profound theoretical backgrounds and complicated mathematical procedures required by IRT and CAT, as well as the large test item bank whose items should have been practically used by more than hundreds of examinees.

2 IRT Calculation for "Lexue 100" System

"Lexue 100" (https://www.lexue100.com), with the Chinese meaning Happy Learning for 100 Percent, is a web-based intelligent tutoring system for school mathematics, developed by Beijing Lexue 100 Online Education Co. Ltd., and equipped with the OLAI (Online Learning Activity Index) model proposed by the authors' team (Jia and Yu 2017; Zhang and Jia 2017). The online intelligent tutoring system "Lexue 100" has been used by more than 100,000 junior high school students in China in the past 7 years. More than 70000 quizzes have been designed for the different versions of mathematics textbooks that are used in different provinces and metropolis in China. Writing quizzes is the main learning activity in this system. Each quiz is composed of a series of gap-filling or single-choice question items with predefined standard answers. As soon as one student submits the trial answer to the system, the trial answer can be compared with the standard answer, and the corresponding quiz score and appropriate feedback are instantly provided to the student. Students are allowed to pass the quiz only if every answer gets right. If the first try of one student is wrong, the student will have to try again until the answer hits the point.

Because the quiz content is the same one for all levels of students, the advantaged students feel writing the quiz boring and time-consuming, while the disadvantaged students feel writing the same quiz challenging (Le et al. 2020). A great need for an adaptive test instead of the stereotyped test is put forward by the students and the school teachers who used this system. In order to meet this need, we designed a computerized adaptive test system based on a large item bank whose items were selected from the quiz set of Lexue 100 system, and then were pre-calibrated and linked with each other.

2.1 The Program for Calibrating, Equating and Linking Parameters in IRT

An open source Java library for psychometric analysis, which is licensed under the Apache License Version 2.0 (https://github.com/meyerjp3/psychometrics), is adopted for calibrating, equating and linking parameters in IRT. The library is comprised of seven modules: (1) core: the base module required by all other module. (2) ctt: classes for CTT including item analysis, test scaling, and reliability estimation. (3) factor: classes for exploratory factor analysis, and for computing polychoric correlations. (4) fmm: classes for multivariate normal mixture models. (5) irt: classes for item response theory, including marginal maximum likelihood estimation (MMLE) for the 1PL, 2PL and 3PL model, scale linking and score equating. (6) nirt: classes for nonparametric IRT, including Ramsay's kernel regression approach for estimating ICCs. (7) optim: a module for optimization.

We mainly used the three modules from this package: core, ctt and irt, and modified them in many parts to calibrate, equate and link the parameters in the 3PL IRT model for "Lexue 100" system.

2.2 Procedure for Parameter Calibrating, Equating and Linking

As shown in our previous research (Le et al. 2020), the student may guess the question answers for many times in order to get the final correct answer. Considering the possibility of students' guessing behavior, we select the 3PL model as the IRT model. If the student obtains the correct answer in the multiple choice or blank-filling question item just with one single trial, we define the response as 1 or correct, otherwise as 0 or false. In addition, we just include the first submission answer to a quiz including some question items of a student as his or her true answer, because the next submissions may be influenced by other factors.

The procedure for parameter calibrating is the concrete implementation of the calibration procedure introduced in Sect. 1.2, and can be described as the following.

(1) For a given test, counts the different examinees who have submitted answers to it. If the examinee number is greater than 100, this test is defined as a valid test whose parameters should be calculated, and go to step (2).
(2) For the valid test, calculate the item responses of all the examinees in the first trial.
(3) With the item responses of the valid test, use the MMLE to calibrate the a (discrimination), b (difficulty), and c (guessing) parameters in 3PL model; then insert the result into a table in the database.
(4) Repeat step (1) through step (3), until the parameters of all question items in all valid tests are computed.

The procedure for parameter equating and linking is the concrete implementation of the calibration procedure introduced in Sect. 1.3, and can be described as the following.

(5) From the IRT parameter table, select the "biggest" test whose examinees are the most as the baseline test, and whose parameters have not been equated and linked. This test is comprised of several quizzes. The already calibrated parameters of the

items in those quizzes do not need to be equated and linked any more, and are inserted in another table for linked parameters.
(6) Search the quizzes in the baseline test in all other tests, get the next “biggest” test with the next most examinees, use the common quiz in the two tests as link quiz, and calculate the equating coefficient with four different methods including the mean-sigma, the mean-mean, the Haebara and the Stocking-Lord methods, and use the bisector method to average four equating coefficients.
(7) Use the mean equating coefficients to link the other quizzes in the next “biggest” test, and insert the equated parameters into the table for linked parameters.
(8) Label the “biggest” test as equated and linked.
(9) Repeat step (5) through step (8) until all tests are checked.

2.3 Result

Each of about 68,600 quizzes among the 70,000 sets has been used by more than 100 students. We analyzed those quizzes, calibrated the original 740,910 IRT parameters for the questions composing the quizzes, recalculated the parameters by using linking algorithms for 194,205 question items.

3 CAT Design and Implementation

Based on the large question banks with pre-calibrated and linked parameters for every question item in it, we designed and implemented a computerized adaptive test system for mathematics in junior high schools as a pilot study. The pilot test is focused on the topic “Pythagorean theorem” in Grade Two of junior high schools in mainland China, which is important but difficult for many ordinary pupils. It is implemented with programming language PHP and embedded in the intelligent tutoring system “Lexue 100”.

The procedure of this CAT is the concrete realization of the procedure introduced in Sect. 1.4, and can be described as the following.

(1) Initialize the test, give the examinee the initial trait value, i.e. the minimum -3.
(2) Search in the linked parameter table for all the quizzed with the topic “Pythagorean theorem”, calculate the information function values for all the quizzes based on the examinee’s current trait value, get the quiz with the maximum information value as the best appropriate one to the examinee.
(3) The best next item is administered and the examinee responds with his or her trial answers.
(4) A new ability estimate is computed based on the responses to all of the administered items.
(5) Steps (2) through (4) are repeated until any one of the ending criteria is met:
(6) the standard error (SE) is less 0.001, what means the examinee’s trait becomes stably measured;
or the trait is equal or greater than 3 for continuous three times, what means the examinee’s ability reached the biggest one;
or the trait is equal or less than -3 for continuous three times, what means the examinee’s ability reached the smallest one.

During writing the CAT, the examinee can get the report about his or her performance. One example of such reports is as the following:

> "*Hello! After 4 rounds of testing, your ability value is −3.000, and the standard deviation is 0.648. In other words, you rank higher than 0.13% of the subjects after ranking their abilities from large to small. If you repeat a similar test, there is a 68% chance that your ability is in the 0.01% to 0.93% position of all subjects.*"

As an example, one difficult question with p value 15.6% in CTT about Pythagorean theorem and rotational congruence is shown in Fig. 1. Drawing auxiliary lines is the key to solve this kind of question.

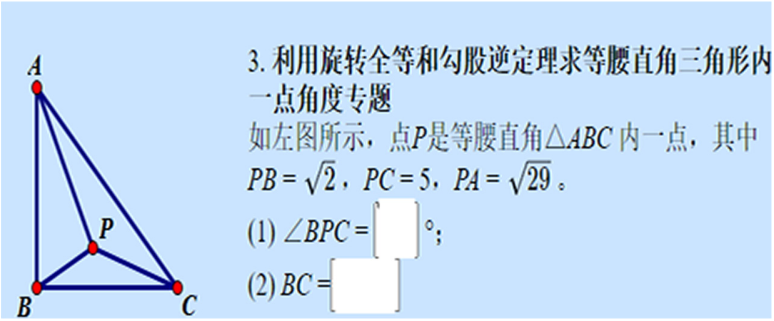


Fig. 1. One difficult question about Pythagorean theorem and rotational congruence

4 Conclusion and Discussion

Rooted in the classical test thoery (CTT), the traditional but ordinary test is one important approach to evaluate the students' mastery of mathematics knowledge and skill. Classical school mathematics tests provide all examinees with the same test content and may be time-consuming or redundant for some examinees. On the contrary, the computerized adaptive test (CAT) stepwise estimates every examinee' response to every question item and provides the examinee with the next corresponding and appropriate question item. In CAT every examinee faces the individual and adaptive questions, whose amount is usually less than that in classical tests, but can receive better and faster assessment result. The prerequisite for a CAT is a large question bank, in which all the questions have been used in previous tests, their answers have been scored, and the parameters for every question such as discrimination, difficulty and guessing, which are defined as three parameters, are pre calibrated according to the item response theory (IRT) with maximum likelihood estimation (MLE). The item parameters from different tests can be equated and linked through common items so that they become comparable and used in the same test.

The online intelligent tutoring system "Lexue 100" has been used by more than 100,000 junior high school students in China in the past seven years. It has more than

70,000 sets of mathematics quizzes designed for different units in the various textbooks. Each of about 68,600 quizzes among the 70,000 sets has been used by more than 100 students. We analyzed those quizzes, calibrated the original 740,910 IRT parameters for the questions composing the quizzes, recalculated the parameters by using linking algorithms for 194,205 question items. Based on this large question banks with pre-calibrated parameters for every question in it, we designed and implemented a computerized adaptive test system for school mathematics. As far as we know, this is the first CAT based on pre-calibrated, equated and linked parameters in a 3PL according to IRT from a large question banks that has been practically used by the students across mainland China.

After pilot assessment, this CAT system will be used by a large number of school students as one part of the intelligent tutoring system. Its effect on students' performance and its benefit on students' quiz time and effort will be investigated and compared with traditional tests.

Acknowledgement. The paper is one achievement of the national educational technology research key project of Central Educational Technology Center China, "Personalized instruction research based on artificial intelligence and big data technology" with the project number 176220009. The research is supported by the project "Lexue 100, Smart Education" of Beijing Lexue 100 Online Education Co., Ltd.

The authors thank all the teachers and students who have participated in the program.

References

Allen, M.J., Yen, W.M.: Introduction to Measurement Theory. Waveland Press, Long Grove (2002)

De Ayala, R.J.: The Theory and Practice of Item Response Theory. Guilford Press, New York (2009)

Eignor, D., Stocking, M, Way, W., Steffen, M.: Case studies in Computer Adaptive Test Design Through Simulation. Educational Testing Service Research Report RR-93-66, Princeton (1993)

Haebara, T.: Equating logistic ability scales by a weighted least squares method. Jpn. Psychol. Res. **22**(3), 144–149 (1980)

Hambleton, R.K., Jones, R.W.: Comparison of classical test theory and item response theory and their applications to test development. Educ. Meas. Issues Pract. **12**(3), 3847 (1993)

Jia, J., Yu, Y.: Online learning activity index (OLAI) and its application for adaptive learning. In: Cheung, S.K.S., Kwok, L.-f., Ma, W.W.K., Lee, L.-K., Yang, H. (eds.) ICBL 2017. LNCS, vol. 10309, pp. 213–224. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-59360-9_19

Kolen, M.J., Brennan, R.L.: Test Equating, Scaling, and Linking: Methods and Practices, 3rd edn. Springer, New York (2014). https://doi.org/10.1007/978-1-4939-0317-7

Le, H., Jia, J., Wong, L.: The analysis of timeout behaviours in online tests. Int. J. Innov. Learn. **27**(3), 310–323 (2020)

Lee, W.C., Ban, J.C.: A comparison of IRT linking procedures. Appl. Meas. Educ. **23**(1), 23–48 (2009)

Lord, F.M., Novick, M.R.: Statistical Theories of Mental Test Scores. Addison-Welsley Publishing, Reading (1968)

Loyd, B.H., Hoover, H.D.: Vertical equating using the Rasch model. J. Educ. Meas. **17**(3), 179–193 (1980)

Marco, G.L.: Item characteristic curve solutions to three intractable testing problems. J. Educ. Meas. **14**(2), 139–160 (1977)

Novick, M.R.: The axioms and principal results of classical test theory. J. Math. Psychol. **3**(1), 1–18 (1966)

Stocking, M.L., Lord, F.M.: Developing a common metric in item response theory. Appl. Psychol. Meas. **7**(2), 201–210 (1983)

Traub, R.: Classical test theory in historical perspective. Educ. Meas. Issues Pract. **16**(4), 8–14 (1997). 9

van der Linden, W.J., Hambleton, R.K.: Handbook of Modern Item Response Theory. Springer, New York (1997). https://doi.org/10.1007/978-1-4757-2691-6

Zhang, B., Jia, J.: Evaluating an intelligent tutoring system for personalized math teaching. In: Proceedings of International Symposium on Educational Technology 2017, pp. 126–130 (2017)

Medical Teachers' Action Research: The Application of Bloom's Taxonomy in Formative Assessments Based on Rain Classroom

Wei Li[1], Hongmei Cai[2,3], Xuesong Yang[2], and Xin Cheng[2](✉)

[1] Academic Affairs Office, Jinan University, Guangzhou 510632, China
674339671@qq.com
[2] Division of Histology and Embryology, Medical College, Jinan University, Guangzhou 510632, China
tchengxin@jnu.edu.cn
[3] Mianhu Middle School, Jieyang 515438, Guangdong, China

Abstract. **Problem** - Both summative and formative assessments are employed in the medical education. Summative assessment (SA) evaluates a student's learning progress and provide concrete grades or pass/fail decisions. Formative assessment (FA) provide the feedback and improve the students' learning behavior. However, it is less known about the impact of FA on student's recognition construction. This study is to exhibit the effects and correlation of FA implementation on students' cognitive performance in final SA at Jinan University School of Medicine. **Intervention** – FA based on teaching management platform - Rain Classroom were performed consecutively in two batches of clinical medicine students. Their scores of final SA constructed on Bloom's taxonomy levels were compared. A questionnaire on the students' learning activities and perceptions was surveyed at the end of each term. **Outcome** - The students' academic performance in 2018 was not statistically significant improved at each Bloom's taxonomy level after the first round of FA implementation, when compared to the students in 2017 who had been evaluated only by final SA. Further methods were taken to strengthen the students' active learning, timely feedback from teachers and peers was supplied on Rain Classroom. After the second round of FA, although exhibiting no statistical difference compared to those of the students in 2018, the performance of the students in 2019 was significantly increased compared to the students in 2017 at the analyzing level. The survey manifested that FA and FA built-in feedback cultivated the students' active learning effectively, reinforced higher-order thinking and problem solving skills which related to higher cognitive levels. Rain Classroom platform facilitates FA implementation. **Conclusion** – It is necessary that incorporating FA into the regular assessment system for optimal educational output. The development of cognitive abilities should be emphasized when designing the assessments methods. Our FA based on Rain Classroom provides an example for online assessments of blended learning.

Keywords: Formative Assessment, FA · Bloom taxonomy · Rain classroom · Medical education

L.-K. Lee et al. (Eds.): ICTE 2020, CCIS 1302, pp. 112–125, 2020.
https://doi.org/10.1007/978-981-33-4594-2_10

1 Existing Problems in Medical Education Assessment

Training the future physicians is a tremendous undertaking that requires empirically supported learning methods and assessment techniques. Assessment is a pivotal aspect of medical education, which is intended to measure students' knowledge, skills, and attitudes, therefore it works as a powerful driver of learning for students [1]. Meanwhile, it represents the quality of students and universities to the outside world. Therefore, teachers should strive to assess students in the most beneficial way for them, and to meet the changing demands of the new generations and of the different assessment stakeholders. However, lack in the evaluation of higher level of thinking skills, affective and psychomotor domains, still are the problems perplexing the educators.

There are two types of assessments generally: summative assessment (SA) and formative assessment (FA). Both approaches exhibit various benefits and limitations. SA has been well established, where teachers measure the achievement of learning goals at the end of a course. As the future doctors, medical students must meet minimal safety standards. Naturally, SA has long been regarded as reliable indicator of the learning outcome and academic performance of medical students. FA, however, provides feedback to students that can be used to improve their learning. Generally, educators think FA as assessment for learning (AfL) while SA is assessment of learning (AoL).

The objectives of FA are to inform the students of required knowledge for the future, to identify student's lack of performance, to assist with improvement in a timely method, and to provide feedback to teachers about how to assist the students. FA exhibits some advantages in medical education. It focuses on specific content, topics or skills, and can be conducted as frequently as possible, it serves the purpose to measure the student's progress over time, and pertains to long-term competencies and expertise; it can improve students' understanding of the subject matter, consolidate the learning and reinforce the learning behavior, help them continuously improving, and motivate the students to learn [2, 3]. However, FA was not enough emphasized at medical schools in China. A course of basic medical science - histology & embryology (HE) is taken for an example. A national wide survey about the current situation of HE teaching was done by our team in 2018, the results showed that 32 of the total 66 surveyed medical schools in mainland China had implemented FA in all programs [4]. The results revealed some challenges of FA during the implementation, including interrater reliability, variability in the predictability of performance scores, difficulties to track long-term development of clinical thinking, and timely feedback, etc.

Nowadays, the progress of smart teaching tools and learning management platforms is changing the formats of assessments. The digitalized protocol of assessments can support and grade the instant response of students, provide a timely feedback to students, and generate reports. The embedded assessments into the learning management platform facilitate FA, making it more effective, sensitive and formative. Nevertheless, the exploration in new fields of assessments is emerging except for the improvement of techniques on assessments. How to develop reliable FA formats using these powerful techniques to cultivate high levels of cognitive ability of the medical students, is a great challenge to the educators. In current study, we performed an action research (AR) of

FA of HE course based on Rain Classroom, a smart teaching tools, to explore and optimize the feasible assessments which was aimed to improve the higher levels of cognitive abilities.

2 Literature Review

2.1 Assessment in Medical Education

Essentially, assessment is a fundamental component of both learning and teaching. Assessment is defined as a process of collecting and evaluating information to measure students' progress [5]. Meanwhile, educators are required to provide means of discrimination among students, also assessment is needed to provide institutional quality assurance to key stakeholders. Therefore, assessment is a powerful tool in the armory of the educators. It is well established that assessment shapes the experience of students and influences their behavior more than any other elements of their education, and deserving careful consideration and practicing [6].

When exploring assessment, there are six key questions which should be addressed: why, what, how, when, where and who [7]. Among them, there are plenty of attempts on how to assess the students. SA has an outward focus serving wider societal aims rather than the students. Placing the students in the center of the examination process, however, assessment should foster and motivate student's learning and equip them with the necessary skills to develop lifelong learning [6]. Cultivation of medical students' clinical ability and personalized development are the goals of medical education. This is what FA serves to promote rather than simply measures learning. FA provides teachers and students with timely and frequent feedback on mastery of course material and learning objectives. In essence, teachers are sampling student learning and providing feedback based on the results to modify instruction and learning experience. Thus students can use feedback to identify the weaknesses of their study. Both SA and FA can be used in the same modality of assessments, and we should be careful not to miss important opportunities to fully develop the potential of a given assessment. By incorporating FA through provision of feedback and stimulation of reflection, students learning outcomes could be exponentially increased, which is exhibited by SA [8].

Next, what blueprint is used to assess the students whether they acquire the necessary knowledge, skills and attitudes? Constructive alignment ensures that the learning objectives are mapped carefully against assessment to ensure maximum possible validity [9]. The four main elements, which need to be aligned, are content, intended learning outcomes, pedagogies and assessment principles. There are many models we can use for an assessment blueprint including Miller's pyramid [10], Bloom's taxonomy [11] and the Dreyfus and Dreyfus [12] spectrum of skills acquisition to name but a few. The common thread to each of these is the evolution of student knowledge from the basic information of a novice to the richness of information of an expert [9].

2.2 Correlation of Assessment with Bloom's Taxonomy in Medical Education

Bloom's Taxonomy is widely used in medical educational research to stratify learning activities into different cognitive levels [13, 14]. It categorizes cognitive activities into six hierarchical levels that range from basic recall to higher educational objectives such as application, analysis and evaluation [11]. Bloom's taxonomy has been adopted as a valuable tool for examining students' learning and classifying examination questions based on the cognitive levels and skills the questions are attempting to assess. Different modified versions of Bloom's taxonomy have been published, to better serve as useful tools for specific areas, to assess student performance and rate educational tasks [15, 16].

Anatomy, a core basic medical subject with high clinical relevancy, keeps appealing the teachers to venture into the research of constructing assessment under the frame of learning taxonomy [17]. Phillips et al. conducted a research in attempt to categorize the different levels of cognitive domains of the students on radiological anatomy using a revised version of Bloom's taxonomy [16]. Zaidi et al. presented a new, subject-specific rating tool for histology MCQs rooted in Bloom's taxonomy, which allowed the teachers to grade histology MCQs reproducibly according to their cognitive levels and to create more challenging examination problems [13]. They reported the incorporation of microscopic images which were new to the learners was often an effective method of elevating histology MCQs to higher Bloom's taxonomy levels.

3 Rain Classroom-Based FA Continuum of HE Course

Two rounds of AR cycles of planning, acting, observing, and reflecting were implemented in this project.

3.1 The Structure of FA and Techniques Evolved of HE Course

At international school of Jinan University (JNU), the total HE course is 108 teaching hours for students of clinical medicine (CM) program, which is divided into 27 two-hour theoretical lectures and 18 three-hour practical sessions. The course is modeled as the preliminary medical knowledge to the first-year medical school students, introducing the microstructure of tissues/organs/systems and their embryonic development. Currently, the teaching staffs at most medical school in China use written tests, that is, SA to assess students' knowledge at the end of the semester. The students are required to take examinations in both HE theory and practice at JNU.

Ideally, we should have a rounded assessment process to address FA and SA simultaneously. At present research, teachers included FA as a part of the grade, which students accumulated toward their final grade in a course. FA was performed in a variety of strategies, the importance of each area of knowledge, skills and attitudes was taken into accounts. General assessment methods used for our FA format were attendance, weekly exercises, drawing/microphotography, 3 mid-semester quizzes, group activities and presentation, and discussion. The results of above components totally occupied 50% of the final marks of the course. The rest 50% was from the final paper test of both theory and practice, which was in the format of SA. The primary SA methods commonly used in undergraduate medical education at JNU are multiple choice questions (MCQ), filling the blanks, short answer questions, essay questions, true or false and case analysis. Each assessment method has a different strength in testing a student's knowledge, skills or attitudes, but a carefully balanced combination was needed to comprehensively reflect the assessment blueprint. The comprehensive assessment was given at the end of the course.

In the present study, FA was performed with the aid of tech-based Rain Classroom, a WeChat mini program developed by MOOC-CN Education (Tsinghua Holdings Co Ltd, Beijing). Rain Classroom has been linked with the teaching management system of JNU since 2017, which is an easy-manipulated plug-in component of PowerPoint supported by WeChat. By scanning the QR-code with WeChat (Fig. 1), the students can view the PowerPoint file on their mobile, interact with teachers by "Danmu", "unclear", or answering the questions on mobile. WeChat is a Chinese online communication platform (Tencent, Shenzhen), which teachers and students are most familiar with. Rain Classroom can promote the comprehensive teaching and learning interaction before, during and after class. During the whole online and offline learning process, Rain Classroom dynamically records all the students' learning behavior, and gives rise to data, providing innovative solutions for panoramically quantifying the teaching and learning activities [18]. Through analyzing and integrating these data, teachers can analyze the teaching effects and learning outcomes quantitatively, making dynamic adjustment of teaching strategies according to the reports.

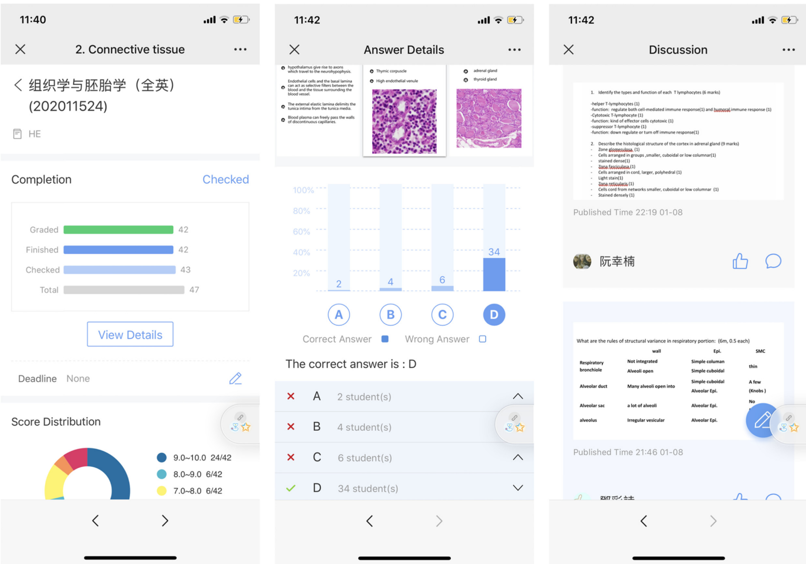


Fig. 1. Snapshot of Rain classroom teaching management platform using for HE course assessments.

3.2 The Construction of SA Based on Bloom's Cognitive Taxonomy

When developing assessment methods, three criteria are emphasized in the HE syllabus, working as the guideline of assessments, "remembering of basic knowledge", "synthesizing and applying", and "analyzing". Based on the understanding of Bloom's objective cognitive pyramid (remember, understand, apply, analyze, evaluate and create), assessment formats were constructed. The "remembering of basic knowledge" are corresponding to "remember" and "understand" of Bloom's taxonomy. "Synthesizing and applying" can be corresponding to "apply". "Analyzing" can be assessed at the level of "analyze" and "evaluate". "Create" is not common in medical science. The "remember" and "understand" categories were collapsed into one category called "knowledge recall", any item in this category was considered a "low" level cognitive item. Items in any of the other categories were considered "high" cognition items [14]. The items were constructed and then evaluated independently by multiple experienced teachers of the course, to categorize according to Bloom's levels of cognition, lastly was reviewed by the department director. These assessments succeed in fulfilling certain educational objectives. Table 1 showed the composition of written examination of HE course at Medical School of JNU during the last 3 years. Examination results were analyzed over a three-year period (the students in 2017 worked as control, while the students in 2018 and 2019 were treatment groups).

Table 1. The composition of final written examination of HE course at Medical School of JNU during the last 3 years.

Terms	Remembering of Basic knowledge (%)	Synthesizing & Applying (%)	Analyzing (%)	Sum
2018–2019 (2)	44	30	26	100%
2018–2019 (1)	44	36	20	100%
2017–2018 (2)	46	30	24	100%
2017–2018 (1)	42	31	27	100%
2016–2017 (2)	50	34	16	100%
2016–2017 (1)	53	34	13	100%

4 Action Research of FA Based on Rain Classroom

4.1 First Round of Action

Student Participants. Since this was a pilot research, 61 medical students were recruited from International School of JNU, which was the total number of medical students of CM program enrolled in 2018 at international school. The pilot project was approved by the Teaching Affairs Unit of JNU during the first term of 2018–2019 academic year. All the students enrolled in the study gave written consents.

Program Development. The usage of Rain Classroom was compulsory to the students during the research, the meaningful data of learning and teaching were acquired and recorded by Rain Classroom, the scores obtained contributed to their final assessment.

In detail, the homework was delivered online after lecture; the 3 mid-semester quizzes were also distributed by Rain Classroom and asked the students to finish within the specific time. The drawing/microstructure pictures were uploaded and scored by peers when completed.

A questionnaire was self-developed to obtain feedback on the FA experience from participating medical students. There were 20 items in the survey, 14 of which used a five-point Likert scale. Students were asked to rank their response based on five points: 1 (strongly disagree), 2 (disagree), 3 (neutral), 4 (agree), and 5 (strongly agree). The survey was piloted by the faculty members of the first author, mainly to ensure the clarity of the questions, and was subsequently revised based on the feedback. The survey was released by Sojump (Ranxing, Changsha, China). All respondents completed the survey online. Their responses were captured anonymously in Sojump which de-identified aggregated information. The results generated through all respondence were collected and reviewed.

All statistical and graphical analyses were performed using the SPSS statistical package, v17.0 (IBM Corp., Armonk, NY). Independent samples *t*-test (2-tailed) was performed to assess the examination scores. Results of the statistical analyses were presented as the mean $\pm$ standard deviation (SD) and the level of statistical significance was accepted to be lower than 0.05 in all analyses. The data obtained from the questionnaires were analyzed using Cronbach's alpha test to determine the internal consistency of the responses. Kendall's tau B test was applied to investigate the relationships among the items of the questionnaire.

Observation and Reflection. The impact of FA of HE course was assessed by analyzing the examination scores achieved by students of CM program in 2018 ($n = 61$) versus students in 2017 ($n = 39$, control). Each student's Bloom's level scores were subtracted from his/her overall final examination paper, the percentage of correct scores against the total scores of each Bloom's level part was calculated as the ratio of the scores. Average performance of the whole class for final examination items assessing each Bloom's level of cognition was calculated. Figure 2A provided the mean performance of knowledge, application, and analyzing HE items for students in 2018 and 2017 respectively. The students in 2018 averaged 75.74 ± 11.46 and the students in 2017 averaged 73.28 ± 12.17 on knowledge items ($P = 0.814$). For application HE items, the students in 2018 averaged 75.57 ± 13.67 and the students in 2017 averaged 73.77 ± 16.30 ($P = 0.915$). Finally, for analyzing HE items the students in 2018 averaged 71.46 ± 17.47 and the studetns in 2017 averaged 65.49 ± 19.27 ($P = 0.156$) (Fig. 2A). The results revealed that the percentage of correct scores of the FA format students was not statistically significantly higher than that of the traditional SA format, at neither of the three Bloom's levels, though it was shown to be improved at each part of Bloom's level.

There were totally 53 of the 61 FA participants responded and completed the questionnaire at the end of the study. The items covered two main areas: the student's attitude and evaluation of FA, the role of Rain Classroom in FA. The Cronbach's alpha for the 14 items in the questionnaire was 0.951, which implied that the survey tool had a good level of internal consistency and reliability. Kendall's tau B was used to test the correlation among the items in the questionnaire. The correlation coefficient was found to range from 0.162 to 0.977, which indicated the items in the survey had positive correlations.

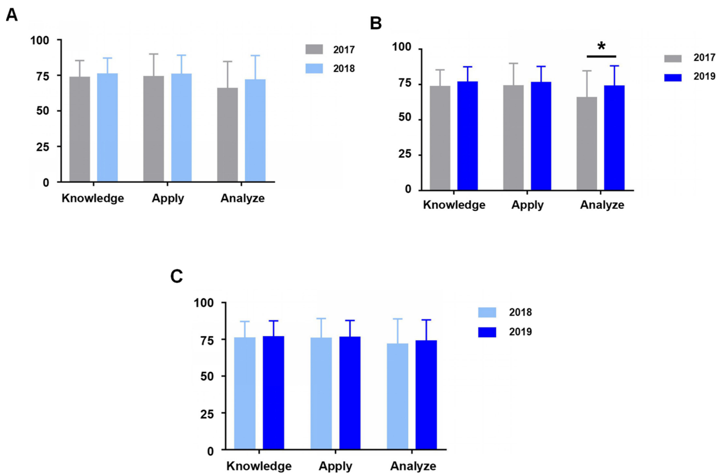


Fig. 2. The comparison of the scores of HE final written examination between students of year 2017 and 2018 (A), year 2017 and 2019 (B), year 2018 and 2019 (C), respectively.

Overall, the questionnaire indicated that majority of participants (67.8%) were positive about the FA (3.77 ± 0.97, Q1), especially the feedback supplied by FA, 81.1% of students believed that FA had helped them to know the gap between what they had known and they should have known (4.19 ± 0.74, Q5). Specifically, 71.7% of the students thought that FA enhanced understanding of the knowledge points (3.87 ± 0.92, Q6), and presentation and reports session promoted further understanding and application of the knowledge points (4.08 ± 0.90, Q7). Generally, the students thought Rain Classroom supported the FA well (3.75 ± 0.92, Q9), including FA supported by Rain Classroom provided feedback from teachers more efficiently (3.94 ± 0.91, Q12), enhanced interaction between teachers and students (3.83 ± 0.91, Q14), and Rain Classroom facilitated self-assessment during learning (3.74 ± 1.02, Q11). However, only around 50% of the participants thought that FA acted as learning guidance to the final examination (3.51 ± 0.95, Q2) and FA exhibited better impacts on future clinical thinking than SA did (3.53 ± 1.08, Q8). The survey revealed that the participants generally acknowledged of and benefited from the FA format, but some students didn't "buy into" FA was positively related to final examination scores and contributed to their future doctor judgements of clinical settings (Table 2). The correlation analyses showed that "FA supported by Rain Classroom enhances interaction between teachers and students" was tightly associated with both "FA enhances understanding of the knowledge points" ($r = 0.935$, $P = 0.00$, n = 53), and "check-in system supported by Rain Classroom contributes to high attendance of the students" ($r = 0.977$, $P = 0.00$, n = 53).

4.2 Second Round of Action

Student Participants. For the second round of the study, 58 CM program students from International School of JNU were recruited in the first term of 2019–2020 academic year.

Table 2. The quantitative statistical descriptive of the questionnaires responded by students in 2018 and 2019 (Mean ± SD).

Items	Students in 2018 (n = 53)	Students in 2019 (n = 55)
1. I prefer FA to SA	3.77 ± 0.97	3.85 ± 0.91
2. FA acts as learning guidance to the final examination	3.51 ± 0.95	3.76 ± 1.02
3. Final written examination is of high reliability of the leaning outcome	3.79 ± 1.01	3.75 ± 0.91
4. Multisourced FA can reflect the learning outcomes much comprehensively	3.85 ± 0.72	3.82 ± 0.92
5. FA helps to know the gap that I have known and I should know	4.19 ± 0.74	3.91 ± 0.91
6. FA enhances understanding of the knowledge points	3.87 ± 0.92	3.78 ± 1.03
7. Presentation and reports session promotes further understanding and application of the knowledge points	4.08 ± 0.90	3.62 ± 1.05
8. FA has a better impact on future clinical thinking than SA	3.53 ± 1.08	3.87 ± 0.92
9. FA supported by Rain Classroom is valuable overall	3.75 ± 0.92	3.65 ± 0.99
10. Check-in system supported by Rain Classroom contributes to high attendance of the students	3.81 ± 0.94	3.40 ± 0.93
11. Rain Classroom facilitates self-assessment during learning	3.74 ± 1.02	3.76 ± 1.02
12. FA supported by Rain Classroom provides feedback from teachers more efficiently	3.94 ± 0.91	4.03 ± 0.82
13. The study efficiency has been improved by using Rain Classroom	3.60 ± 1.06	3.80 ± 0.76
14. FA supported by Rain Classroom enhances interaction between teachers and students	3.83 ± 0.91	3.69 ± 0.96

The overall syllabus, the course contents, the student evaluation and grading policy, and the faculty teaching staff of the course remained unchanged compared to the first round.

Study Design. Targeted at problems appeared in the first round, design of FA was adjusted in second round of action research process.

Except for the FA strategies used in the first round, we assigned three essay questions to the students at different time windows of learning. The first question was "using an example to tell the structures and functions of one tissue type", which was distributed to the students at the end of general histology (the contents on different tissue types in human body). The second question was "using an example to tell the structures and functions of an organ, and further interpret how the tissue types you talked in last homework works in the microanatomic architecture of the organ", this homework was assigned at the end of special histology (the contents on different organs and systems in human body). The third question was "to tell the embryonic origin and development of the organ and tissue you discussed in last two pieces of homework", which was asked when completing embryology (the science of human development). These three questions are the main plots of HE course, one built upon other, which needs synthesizing the knowledge periodically and applying into detailed examples. The questions encourage and promote deep learning of the course. The students were required to complete the essay homework at the discussion area of Rain Classroom, then read and commented at least three other homeworks.

Observation and Reflection. Figures 2B–C provided the mean performance for knowledge, application, and analyzing HE items for students in 2019 via those in 2017, and students in 2019 via those in 2018. The students in 2019 averaged 76.43 ± 11.19 on knowledge items, which showed no statistical significance when compared with students in 2017 ($P = 0.632$) and 2018 ($P = 0.989$). For application HE items, students in 2019 averaged 76.10 ± 11.80, exhibiting no significant difference when compared with both students in 2017 ($P = 0.811$) and 2018 ($P = 0.995$). However, the students in 2019 averaged 73.71 ± 14.57 on analyzing HE items, showing significance compared with those in 2017 ($P = 0.02$), while no difference compared to the students in 2018 ($P = 0.746$, Fig. 2B–C).

The same questionnaire was used for the feedback of Rain Classroom based assessments, and totally 55 participants completed the survey (Table 2). The Cronbach's alpha for the 14 items in the questionnaire was 0.856. The correlation coefficient was found to range from 0.292 to 0.981 among the first 8 items, and range from 0.334 to 0.940 among items 9–14, which indicated the items in the survey have a positive correlation. The correlation analyses showed that "the students prefer FA to SA" was tightly associated with both "FA has a better impact on future clinical thinking than SA" ($r = 0.981$, $P = 0.00$, n = 55), and "multisourced FA can reflect the learning outcomes much comprehensively" ($r = 0.959$, $P = 0.00$, n = 55).

5 Discussion

Assessment is widely used in any subject, to ensure that students reach a specified standard, to be called competent in that subject. Moreover, assessment drives learning. Therefore, the desired career-directed learning can be performed, if the correct type of learning opportunities and assessment are adopted. This study showed that for a first-year medical school HE curriculum, the FA formats improved students' ability to analyze material in the final examination.

The analysis of the questionnaire in this study helped explain why FA promoted the examination performance. From the survey, it can be recognized that FA helped the students to know the gap that they had known and they should have known. Providing students with opportunities of self-correct and self-study is a key element of FA, this process was facilitated by Rain Classroom in this study. When students see the gap between what they thought they knew and what they actually knew, the motivation of learning actually increases. That is effective feedback built-in FA processing, which is quite valuable during learning. Feedback lets students know how they are doing and provide opportunities to adjust and perfect their efforts. Several publication mentioned FA participation was a better predictor of final outcome, even than success in FA, which supporting feedback as the key role of FA, also implying both positive and negative feedback are crucial in learning and development [8]. From the perspectives of educators, feedback from students is collected real time when carrying out FA, used to modify course, and improve teachers' instructive strategy. Therefore, feedback directed learning and teaching strategies provides opportunities of active learning.

Research demonstrates that students engaged in active learning exercises have improved higher-order thinking and problem solving skills [19], which is also verified in this study. Bloom's Taxonomy is widely used in medical educational research, to stratify learning activities into different cognitive levels, and has been adopted as a valuable tool for examining students' learning and to classify examination questions based on the cognitive levels and skills the questions are attempting to assess. In current study, analyzing ability - a higher learning objective was enhanced greatly during FA format teaching strategy. Analysis is where the skills that are commonly thought as critical thinking enters, including the component skills of analyzing arguments, making inferences using inductive or deductive reasoning, judging or evaluating, and making decisions or solving problems [20]. Critical thinking involves both cognitive skills and dispositions. The clinical thinking of medical students is specialized critical thinking, which is required to reach a correct diagnosis and give rise to treatment strategies within a short time and at the lowest cost. This kind of thinking needs specific training during their learning, especially the training of high level cognition.

Most students perceived that Rain Classroom was an effective and attractive tool to support the feedback of FA. Rain Classroom facilitates students' comparisons of their own and peers' progress, enables a social connectivity, and creates learning environment of feedback. This was further reflected by the high correlations between "FA supported by Rain Classroom enhances interaction between teachers and students" and "FA enhances understanding of the knowledge points". As teaching management platform, the Rain classroom is also functionalized in supervision and inspection, the students bought into the "check-in system supported by Rain Classroom", and guaranteed high attendance. Studies have highlighted that technology sponsors new ways of providing feedback [21]. Providing automatic feedback and repetitive opportunities using online quizzes helps students learn through diagnosing their own mistakes [22]. It is well accepted that medicine is regarded as an experienced science. The nature of the online course provides more opportunities to acquire feedback from repetitive exercise.

One limitation of this study was the learning habits of the students were not surveyed and evaluated, which should be given more consideration and analysis, to optimize

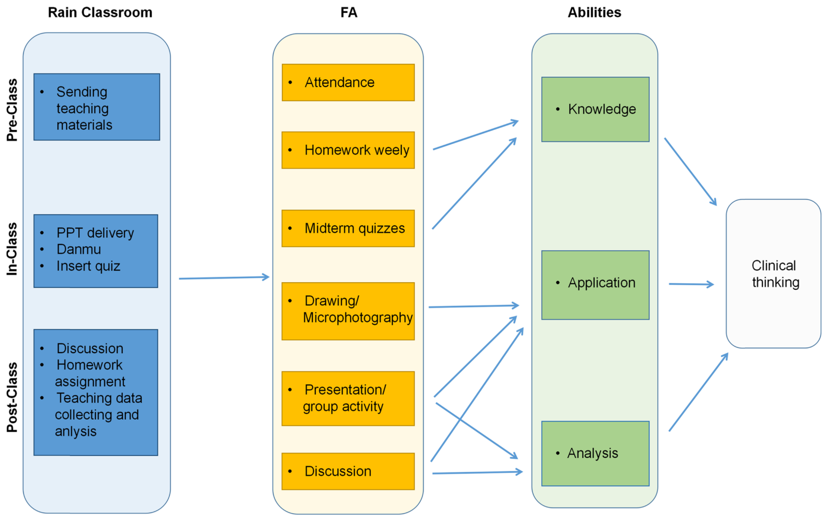


Fig. 3. The relationship of Rain classroom supported learning activities and formative assessments compositions, and their contribution to the different levels of cognitive abilities.

the effect of feedback built-in FA. Secondly, besides concentrating on the low level cognitive thinking, an ideal assessment of basic medical sciences should also be able to promote cognitive application of basic medical knowledge, cognitive synthesis of separated medical concept, and general development of professionalism, which might be constrained if built by a single cognitive framework. We recommend the future project aims at integrating different cognitive taxonomy framework for students' assessments.

6 Conclusion

Effectiveness of using FA in the teaching and learning process of medical sciences has been stressed greatly. The purpose of the current study was to determine the contribution of different formative assessment methods based on Rain Classroom, to the improvement of cognitive abilities at different Blooms' taxonomy levels. The results showed that using technology based on teaching management tools facilitates FA and FA built-in feedback. When the teachers create an effective learning environment that cultivates higher cognitive levels, FA improves students' achievements (Fig. 3). Our FA strategies based on Rain Classroom practicing can provide an example for online assessments, especially for the abrupt transition from face to face teaching to online teaching under the outbreak of COVID-19 pandemics.

Funding. This research was supported by the Pedagogical Reform of High Education at Guangdong Province (2019-60); the research projects of pedagogical reform at Jinan University (JG2019012); and the grant of China Association of Higher Education (2020JXYB08).

References

1. Wormald, B.W., Schoeman, S., et al.: Assessment drives learning: an unavoidable truth? Anat. Sci. Educ. **2**(5), 199–204 (2009)
2. Muhd, A.I., Anisa, A., et al.: Using Kahoot! As a formative assessment tool in medical education: a phenomenological study. BMC Med. Educ. **19**(1), 230 (2019)
3. Sateesh, B.A., Yogesh, A., et al.: Implementation of formative assessment and its effectiveness in undergraduate medical education: an experience at a Caribbean Medical School. MedEdPublish (2018)
4. Cheng, X., Chan, L.K., Li, H., et al.: Histology and embryology education in China: the current situation and changes over the past 20 years. Anat. Sci. Educ. (2020)
5. Holmboe, E.S., Sherbino, J., Long, D.M., et al.: The role of assessment in competency-based medical education. Med. Teach. **32**(8), 676–682 (2010)
6. Epstein, R.M.: Assessment in medical education. N. Engl. J. Med. **356**(4), 387–396 (2007)
7. Ann, F.H., Dermot, O.F.: Assessment in medical education; what are we trying to achieve? Int. J. High. Educ. **4**(2), 139–144 (2015)
8. María, T.C., Eva, B., Caseras, X., et al.: Formative assessment and academic achievement in pre-graduate students of health sciences. Adv. Health Sci. Educ. Theory Pract. **14**(1), 61–67 (2009)
9. Khan, K., Ramachandran, S.: Conceptual framework for performance assessment: competency, competence and performance in the context of assessments in healthcare–deciphering the terminology. Med. Teach. **34**(11), 920–928 (2012)
10. Miller, G.E.: The assessment of clinical skills/competence/performance. Acad. Med. **65**(9 Suppl.), S63–S67 (1990)
11. Krathwohl, D.R., et al.: Taxonomy of educational objectives: the classification of educational goals. In: Handbook I: Cognitive domain, Longmans (1956)
12. Carraccio, C.L., et al.: From the educational bench to the clinical bedside: translating the Dreyfus developmental model to the learning of clinical skills. Acad. Med. **83**(8), 761–767 (2008)
13. Nikki, B.Z., Charles, H., et al.: Climbing Bloom's taxonomy pyramid: lessons from a graduate histology course. Anat. Sci. Educ. **10**(5), 456–464 (2017)
14. Morton, D.A., Colbert-Getz, J.M.: Measuring the impact of the flipped anatomy classroom: the importance of categorizing an assessment by Bloom's taxonomy. Anat. Sci. Educ. **10**(2), 170–175 (2017)
15. Thompson, A.R., O'Loughlin, V.D.: The Blooming Anatomy Tool (BAT): a discipline-specific rubric for utilizing Bloom's taxonomy in the design and evaluation of assessments in the anatomical sciences. Anat. Sci Educ. **8**(6), 493–501 (2015)
16. Phillips, A.W., Smith, S.G., et al.: Driving deeper learning by assessment: an adaptation of the Revised Bloom's Taxonomy for medical imaging in gross anatomy. Acad. Radiol. **20**(6), 784–789 (2013)
17. Hadie, S.N.H., Abdul, H., et al.: Creating an engaging and stimulating anatomy lecture environment using the cognitive load theory-based lecture model: students' experiences. J Taibah Univ Med Sci. **13**(2), 162–172 (2018)
18. Ren, Q., Zhang, J., et al.: Research and practice of flipped classroom based on rain classroom under the background of big data. In: Striełkowski, W., Cheng, J. (eds.) International Conference on Management, Economics, Education, Arts and Humanities (MEEAH 2018), pp. 120–124. Atlantis Press (2018)
19. Stockwell, B.R., Stockwell, M.S., et al.: Blended learning improves science education. Cell **162**(5), 933–936 (2015)

20. Adams, N.E.: Bloom's taxonomy of cognitive learning objectives. J. Med. Libr. Assoc. **103**(3), 152–153 (2015)
21. Ismail, E., Abdulghani, A., et al.: Using technology for formative assessment to improve students' learning. Turk. Online J. Educ. Technol. **17**(2), 182–188 (2018)
22. Baeiter, O., Enstroem, E., et al.: The effect of short formative diagnostic web quizzes with minimal feedback. Comput. Educ. **60**(1), 234–242 (2013)

Automatic Generation of Matching Rules for Programming Exercise Assessment

Tak-Lam Wong[1](✉), Chung Keung Poon[2], Chung Man Tang[3], Yuen Tak Yu[4], and Victor Chung Sing Lee[4]

[1] Department of Computing Studies and Information Systems, Douglas College, New Westminster, British Columbia, Canada
tlwong@ieee.org

[2] Department of Computing, The Hang Seng University of Hong Kong, Shatin, Hong Kong SAR
ckpoon@hsu.edu.hk

[3] School of Science and Technology, The Open University of Hong Kong, Kowloon, Hong Kong SAR
cmtang@ouhk.edu.hk

[4] Department of Computer Science, City University of Hong Kong, Kowloon, Hong Kong SAR
{csytyu,csvlee}@cityu.edu.hk

Abstract. Automatic programming exercise assessment aims at determining the correctness of the attempts of programming exercises submitted by students. Automation allows students to receive instant and customized feedback which are important to enhance the learning of novice students. Educators can benefit from saving time and effort in marking students' attempts, making teaching large, online classes or Massive Open Online Courses (MOOC) possible and effective. Recently, we modelled program outputs using Hierarchical Program Output Structure (HiPOS), which allows instructors to design matching rules to determine correct or partially correct programs depending on the teaching and learning needs. This paper extends our previous work by automating the matching rule construction process through developing a machine learning method for generalizing program outputs from students' attempts. To achieve this, our approach firstly employs natural language processing techniques to create a HiPOS from a set of students' program outputs. A greedy algorithm is then applied to generalize the HiPOS and create the associated matching rules. We conducted a case study to illustrate how to apply our proposed method in automated programming exercise assessment and demonstrated the usefulness and effectiveness of our approach.

Keywords: APAS · Automated programming exercise assessment · Hierarchical program output structure · HiPOS

1 Introduction

Computer programming is an essential competency in many different disciplines such as computer science, data science and engineering because of its practical applicability and the training of computational thinking. However, learning programming is no

L.-K. Lee et al. (Eds.): ICTE 2020, CCIS 1302, pp. 126–135, 2020.
https://doi.org/10.1007/978-981-33-4594-2_11

easy task because it requires the acquisition of different forms of knowledge including problem formulation, language syntax, logical thinking, solution implementation, debugging and testing, etc. Many novice learners find computer programming complicated and abstract, and become lack of motivation (Law et al. 2010). Different learning and teaching methods have been proposed and investigated to enhance programming learning. For example, many educators have designed different gamification strategies to help learners learn programming via playing video games (Lindberg et al. 2019). Practice makes perfect. While these methods were proved successful in motivating novice learners and assisting them in acquiring some basic concepts, a key success factor is to practice as many different types of programming exercises as possible. Traditionally, instructors design assignments and learners attempt the assignments in order to learn the topics of interest. Instructors are responsible to judge the correctness of the program after checking students' submission. Obviously, learners need to wait for the marking and feedback by their instructors after completing the assignments. Research has already shown the importance of timely feedback to learners (Dawson et al. 2019; Lizzio and Wilson 2008). Delayed response by instructors can adversely affect the learning experience and learning outcomes. On the other hand, due to the limited resources, instructors are not able to create a wide variety of assignments and mark all the attempts submitted by a large class of students, not to mention the provision of detailed feedback, even though some learners may wish to practise more exercises (Henderson et al. 2019).

The limitation of traditional approach to learning and teaching of programming leads to the need of intelligent tutoring systems (ITS), which allow learners to practice programming exercises at any time. In an ITS, instructors can create a number of programming exercises for the learners to attempt. At the same time, the test cases together with the expected outputs will also be imported into the ITS. After a learner submits an attempt, the ITS will verify the *admissibility* of the submitted program automatically by comparing the actual output of the submitted program and the expected output. A program or program output is said to be *admissible* if it is considered correct by the instructor even if it deviates from an expected correct program or program output, respectively (Poon et al. 2016). Learners will be notified the results soon after the submission so that they can know their progress and level of understanding on the topic, and then decide the next step of the learning path. The instant response and the availability of large amount of exercises allow learners to do a lot of practice according to their pace and effort. At the same time, ITS can save the time and effort of instructors in marking the attempts. This also makes organizing a very large class like Massive Open Online Courses (MOOCS) manageable and possible.

A recent survey study reveals that there exist different types of ITS for programming education, targeting different programming languages and offering different features (Crow et al. 2018). A core subsystem of these ITSs is the Automated Programming Assessment System (APAS), which is to run the learner's program codes to produce the output, and then compare the actual output with the expected output to judge the admissibility of the program. However, most existing APASs can only conduct simple or rigid comparison between program outputs like what those judging systems[1] of programming

[1] Example of judging systems of programming contests include PC^2 (https://pc2.ecs.csus.edu/) and Kattis (https://open.kattis.com/).

contests do. They only accept one single exact output or end-space-trimmed output for each test case as an admissible answer. An output variant which includes, for example, an extra space in the middle of the output, can be treated as incorrect, though it may be judged as correct or admissible by instructors in the merit of learning and teaching. While it is objective, fair and appropriate to impose strict output format requirement in programming contests, in which the contestants are likely to be competent programmers, such a requirement incorporated in an APAS cannot provide any flexibility for novice learners and meet the pedagogical needs of instructors. A novice learner tends to write programs which may produce outputs that do not exactly match the instructors' expected program outputs. When such programs are repeatedly deemed "incorrect" and rejected by an APAS even though they are actually admissible, the learner will very likely become highly discouraged and demotivated, ultimately seriously affecting the learning outcomes.

Recently, we developed a model called Hierarchical Program Output Structure (HiPOS) to address the problem of multiple admissible output variants to support programming education (Poon et al. 2016). HiPOS is a tree-like structure, which can be associated with different levels of matching rules. HiPOS together with the matching rules can correctly identify a wide variety of admissible output variants based on the pedagogical needs of instructors. When creating an exercise as well as its test cases in APAS, instructors create these matching rules to specify the possible admissible output variants. HiPOS is shown to be flexible to handle more complex programming outputs including natural language text, semi-structured text, or tabular formatted outputs. While the manual effort of creating such matching rules is not huge, we further reduce such effort through extending our previous work. By using a machine learning approach, we developed a method to automatically learn the matching rules from the training examples, which can be a collection of previous students' attempts in the exercise. Essentially, we develop a greedy algorithm to generalize the matching rules from a set of admissible program output variants. Through this automatic generation of matching rules, instructors can save the effort and time in creating the rules manually. On the other hand, these matching rules are able to cover more admissible output variants (that is, reducing false negative cases) since they are learned from previous learners' attempt instead of being created by instructors.

The organization of the rest of the paper is as follows. We firstly present the recent and related work to APAS in Sect. 2. In Sect. 3, we provide some background about HiPOS and describe our learning algorithm to discover those matching rules for HiPOS from a set of training examples. It is followed by a case study on applying our learning algorithm to automatically create HiPOS and the associated matching rules from training examples. The learned HiPOS is then utilized to determine the admissibility of the previously unseen attempts by students. Finally, Sect. 5 concludes the paper.

2 Related Work

Wasik et al. (2018) conducted a comprehensive survey of online judging systems and discussed their applications. Since the primary purpose of these systems are for competition among competent programmers or contestants, they usually adopt very strict program output comparison methods and, therefore, are not desirable for use in learning and teaching for novice learners. Our work described in this paper is more concerned with APASs which help instructors to assess the programs submitted by novice learners. Some existing program output comparison methods rely on instructors' effort in constructing regular expression so as to assess program output variants with certain flexibility (Queirós and Leal 2012). A *program output variant*, or simply an *output variant*, is a program output which deviates from the instructor's expected output and may be admissible or inadmissible. Pieterse (2013) proposed a system which compares program output variants against the expected program output using either exact matching or regular expressions. Mekterović et al. (2020) conducted a survey on APAS and developed a system, Edgar, which supports regular expressions. The regular expressions need to be specified by the instructor and, therefore, it is time consuming, error-prone and requires a high level of expertise. Coderunner is a recent system which can assess submitted programs in multiple languages (Lobb and Harlow 2016). In addition to exact matching, instructors may develop a validator, which is essentially another program in Python, to assess the admissibility of the output variants. However, writing and debugging a program as validator is also time consuming and error-prone.

Different from the aforementioned approaches, Token Pattern Approach (TPA) employs *token patterns* to assess the program output variants (Tang et al. 2009). For each token in the expected output, instructors can specify customized rules for matching with the corresponding actual output token, and the collection of all such tokens with the associated matching rules forms a token pattern. Experiments have been performed by incorporating TPA into an existing APAS for evaluation and demonstrated promising results. However, the performance and the applicability of TPA is limited by the need of a separate effort for the instructor to create token patterns, including the matching rules. To address this limitation, we proposed Hierarchical Program Output Structure (HiPOS), which is more robust and flexible to handle various types of program output variants (Poon et al. 2016). A description of HiPOS will be provided in the next section.

3 Methodology

We have developed the Hierarchical Program Output Structure (HiPOS) to model the program output for automated programming assessment and reported in (Poon et al. 2016; Poon et al. 2018). We provide a brief description of HiPOS in this section, followed by an explanation of our approach in learning matching rules for HiPOS. HiPOS is a tree-like structure consisting of nodes and edges. Every node has one and only one parent node connected by an edge, except for a special node known as root which does not have any parent node. A leaf node of a HiPOS refers to a token of the program output, while an internal node represents a block of tokens. A node of HiPOS may be associated with a set of matching rules, which are used for determining if a program output variant is admissible. There are four levels of matching rules: problem-level, block-level, token-level, and test-case-level for discovering different types of possible errors in the submitted program. Problem-level matching rules are designed for judging the basic criteria of correctly completing the exercise. They are not specific to any test case, but are applicable to the entire HiPOS and all output variants of the exercise.

Block-level matching rules are applicable to a block, that is, a subtree rooted at an internal node of a HiPOS of an output variant. Token-level matching rules are applicable to a single token. They are associated with a leaf in HiPOS. Test-case-level matching rules are designed for a particular test case, usually for extreme cases only. Consider the programming exercise, **Ex.1**, as follows.

> **Ex.1**: *Write a program that asks the user to input a date (d/m) in 2005, and print the weekday of the date.*

Given the test case with the input "**6/4**" (April 6th), the expected output set by the instructor is "**6/4/2005 is Wednesday**". However, some other program output variants like "**6/4 is Wednesday**" and "**6/4/2005 is a Wednesday.**" may also be considered as admissible. Figure 1 shows the HiPOS which is automatically created using a natural language parser and the possible matching rules created by the instructor. It can be employed to assess the admissibility of the program output variants by students' submitted programs.

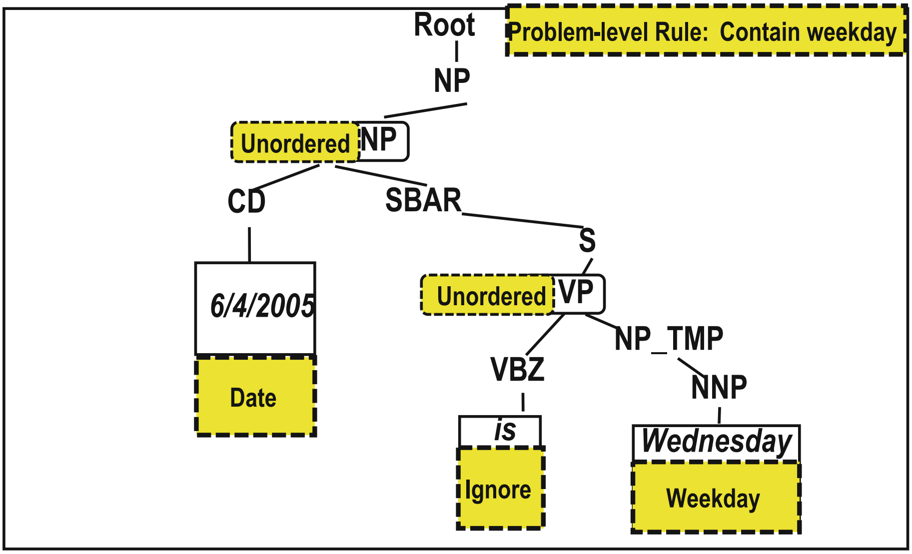


Fig. 1. An example of HiPOS with matching rules for the output "6/4/2005 is Wednesday".

Our previous work shows that HiPOS is effective in identifying different types of admissible program output variants automatically. One characteristic of HiPOS is that the matching rules are created by instructors to accommodate their pedagogical needs. However, manual work is required to construct these matching rules. Manual construction may also lead to unintended errors in the matching rules, in turn affecting the accuracy of HiPOS in determining the admissibility of output variants and the learning experience of students. To address this problem, we have developed a greedy approach to learning the set of matching rules from a set of training examples.

Input: A set of admissible program output, ***POS***
A set of inadmissible program output, ***NEG***

Output: HiPOS of the expected program output, *H*.
The set of matching rules associated with HiPOS, *M*.

1. Randomly select a seed from ***POS*** and create a HiPOS, denoted as H^0.
2. $n \leftarrow 0, M^0 \leftarrow \{\}$
3. Until no accuracy improvement:
  3.1 $n \leftarrow n + 1$
  3.2 For each leaf node, $node_i$, in H^0:
    3.2.1 Create candidate token-level matching rule by generalizing the token of $node_i$, according to the token generation hierarchy, R_{cand}
    3.2.2 Create M_i^n by adding R_{cand} to M^{n-1}
    3.2.3 Test H^0 and M_i^n on $\boldsymbol{POS} \cup \boldsymbol{NEG}$
  3.3 Set M^n to M_i^n such that M_i^n has the best accuracy on $\boldsymbol{POS} \cup \boldsymbol{NEG}$
4. Until no accuracy improvement:
  4.1 $n \leftarrow n + 1$
  4.2 For each internal node, $node_j$ in H^0:
    4.2.1 Create candidate block-level matching rule by adding unordered or semantic similarity match rule, R_{cand}
    4.2.2 Create M_j^n by adding R_{cand} to M^{n-1}
    4.2.3 Test H^0 and M_i^n on $\boldsymbol{POS} \cup \boldsymbol{NEG}$
  4.3 Set M^n to M_i^n such that M_i^n has the best accuracy on $\boldsymbol{POS} \cup \boldsymbol{NEG}$
5. Return H^0, M^n

Fig. 2. Matching rules learning algorithm

Figure 2 shows our proposed learning algorithm for generating the set of matching rules. In our algorithm, we firstly select a random admissible program output variant from the set of admissible program output variants denoted by ***POS*** to generate the candidate HiPOS, H^0. We adopted natural language toolkit[2] to generate the HiPOS. Since H^0 is not associated with any matching rule, it is the most restricted HiPOS. After that, the algorithm aims at generating the token-level matching rules by generalizing each token according to the token generation hierarchy depicted in Fig. 3. In this hierarchy, higher level nodes are more general than the lower nodes. The generalization will be carried out in a bottom-up manner according to the token generalization hierarchy. For example, given the token "**6/4**", it can be generalized to matching rules in the order of "**TimeDayMonthYear**", "**DayMonthYear**", "**DayMonth**" and "**Ignore**" in our learning algorithm. In each iteration (Step 2), only one token will be generalized and the one resulting in the best accuracy will be added to the set of matching rules, M^n. Therefore, Step 2 can create the matching rules associated with relevant nodes of HiPOS by incorporating one token-level matching rule at a time, until there is no more improvement in accuracy. Similarly, Step 3 aims at adding one block-level matching rule at a time until there is no more accuracy improvement. Subsequently, Step 4 returns the HiPOS associated with the set of matching rules which achieves the best accuracy.

[2] Natural Language Toolkit: https://www.nltk.org/.

Problem-level and test-case-level rules are not considered in our algorithm because they can be created by Instructor easily. The learned HiPOS together with the matching rules can then be utilized to determine the admissibility of program output variants. If necessary or desired, instructors can further revise the resulting set of matching rules to better accommodate their pedagogical need.

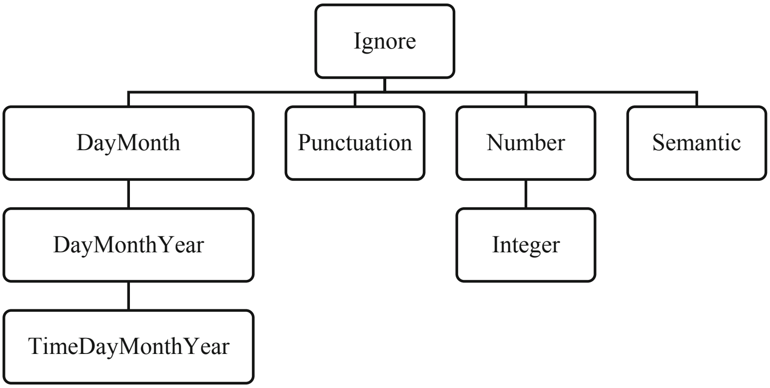


Fig. 3. Token generalization hierarchy for matching rule generation

Our matching rule learning algorithm adopts a greedy approach to achieve a reasonably small training time. The overall time complexity of our algorithm is $O(NLH)$ where N, L and H refer to the number of training examples, the number of leaves and the height of the token generation hierarchy respectively. On the other hand, Steps 3.2 and 4.2 can be run in parallel and hence distributed training across multiple processors can be employed to speed-up the training process.

4 Case Study

We implemented our matching rule learning algorithm and applied our method to real-world data to evaluate our method. Referring to the programming exercise depicted in **Ex.1** and the expected program output, we investigated the programs submitted by 153 students to attempt this question. Each program was tested by 84 different test cases and produced ($153 \times 84 =$) 12852 program outputs for evaluation. We then removed those outputs which are exactly the same as the expected output after end-space trimmed. Among the remaining 4655 program output variants, 2791 of them are admissible while 1864 of them are inadmissible. These program output variants were assessed manually and labelled as admissible or inadmissible. We then randomly divided these program output variants into two equal sets: a training set and a test set.

We applied our matching rule learning algorithm in the training set to discover the matching rules and generate the HiPOS. The learned HiPOS together with the matching rules were then applied to the test set to predict the admissibility of each program output variant. In reality, instructors can further modify the matching rules to suit their pedagogical needs if they desire. However, in our case study, we did not make any modification in order to evaluate the performance of the learning algorithm. The predicated admissibility was compared with the manually labelled admissibility for evaluation. We first computed the accuracy, that is, the total number of correct prediction divided by the total number of test data. We also computed the precision, recall and $F1$-measure for the admissible program output variants. Precision is defined as the number of correctly identified admissible program output variants divided by the total number of predicted admissible program output variants. Recall is defined as the number of correctly identified admissible program output variants divided by the actual number of admissible program output variants. $F1$-measure is defined as the harmonic mean of precision and recall. To demonstrate the effectiveness of our learning algorithm, we apply our previous work, namely, TPA, to assess the admissibility of test set for comparison (Tang et al. 2010). Table 1 shows the experimental results of these two methods. Our learning algorithm achieves an accuracy, precision, recall and $F1$-measure of 85.7%, 92.5%, 82.9% and 87.4% in the test set respectively. TPA obtains an accuracy, precision, recall and $F1$-measure of 83.1%, 100%, 72.1% and 83.8% respectively. The accuracy and $F1$-measure show that our learning algorithm achieves a better performance compared with TPA. A higher recall shows that the learned matching rules are more effective in identifying admissible program output variants than TPA. On the other hand, the instructor can easily improve the precision by manually modifying the learned matching rules to eliminate the false positives.

Table 1. Experimental results of our approach and TPA

	Accuracy	Precision	Recall	*F1*-measure
Our approach	85.7%	92.5%	82.9%	87.4%
TPA	83.1%	100%	72.1%	83.8%

Our learning algorithm is promising in discovering matching rules from the training examples. Practically, the token generalization hierarchy can be modified to accommodate different types of program output variants. For example, program output involving HTML elements can be handled by token generation hierarchy incorporated with different types of HTML elements such as headers, tables, lists, etc. While our learning algorithm is robust and effective, there exists certain limitation. Our algorithm depicted in Fig. 2 is a greedy approach and one matching rule will be added to the resulting HiPOS in each iteration. The major rationale for adopting a greedy strategy is to reduce training time and complexity. Nevertheless, it is possible that it will result in a sub-optimal solution. Take an expected output which contains a pair of parentheses (i.e., "(" and ")") as an example. Supposing that the program output variant is admissible if this pair of parenthesis is replaced with a pair of braces (i.e., "{" and "}"), but inadmissible if only one of the parentheses is replaced with a brace. Our learning algorithm is not able to generalize these two tokens at the same time. In any case, instructor may still revise the resulting HiPOS and the matching rules to accommodate this limitation.

5 Conclusions

We have developed a machine learning algorithm to learn the matching rules for HiPOS from a set of training examples of program output variants. The greedy algorithm we proposed can automatically create token-level and block-level matching rules by generalizing the token and the block of tokens respectively. We presented a case study and compared the performance with the existing work TPA. The results demonstrated that our learning algorithm achieves a better performance and reduces the human effort in constructing the matching rules. We intend to extend our work in different directions. The first possible extension is to integrate our method to an APAS and apply it in daily teaching and learning. We aim at investigating the perception to HiPOS by students and studying the feedback from instructors. Another possible direction is to develop a pedagogical framework which integrates HiPOS and the learning algorithm to enhance programming education to novice learners.

Acknowledgement. The work described in this paper is fully supported by a grant from the Research Grants Council of the Hong Kong Special Administrative Region, China (Project UGC/FDS11(14)/E02/15).

References

Crow, T., Luxton-Reilly, A., Wuensche, B.: Intelligent tutoring systems for programming education: a systematic review. In: Proceedings of the 20th Australasian Computing Education Conference, pp. 53–62 (2018)

Dawson, P., et al.: What makes for effective feedback: staff and student perspectives. Assess. Eval. High. Educ. **44**(1), 25–36 (2019)

Henderson, M., Ryan, T., Phillips, M.: The challenges of feedback in higher education. Assess. Eval. High. Educ. **44**(8), 1237–1252 (2019)

Law, K.M., Lee, V.C., Yu, Y.T.: Learning motivation in e-learning facilitated computer programming courses. Comput. Educ. **55**(1), 218–228 (2010)

Lindberg, R.S., Laine, T.H., Haaranen, L.: Gamifying programming education in K-12: a review of programming curricula in seven countries and programming games. Br. J. Edu. Technol. **50**(4), 1979–1995 (2019)

Lizzio, A., Wilson, K.: Feedback on assessment: students' perceptions of quality and effectiveness. Assess. Eval. High. Educ. **33**(3), 263–275 (2008)

Lobb, R., Harlow, J.: Coderunner: a tool for assessing computer programming skills. ACM Inroads **7**(1), 47–51 (2016)

Mekterović, I., Brkić, L., Milašinović, B., Baranović, M.: Building a comprehensive automated programming assessment system. IEEE Access **8**, 81154–81172 (2020)

Pieterse, V.: Automated assessment of programming assignments. In: Proceedings of the Computer Science Education Research Conference (CSERC), pp. 4–5 (2013)

Poon, C.K., Wong, T.L., Yu, Y.T., Lee, V.C., Tang, C.M.: Toward more robust automatic analysis of student program outputs for assessment and learning. In: Proceedings of the 2016 IEEE 40th Annual Computer Software and Applications Conference (COMPSAC), vol. 1, pp. 780–785 (2016)

Poon, C.K., Wong, T.L., Tang, C.M., Li, J.K.L., Yu, Y.T., Lee, V.C.S.: Automatic assessment via intelligent analysis of students' program output patterns. In: Proceedings of the International Conference on Blended Learning, pp. 238–250 (2018)

Queirós, R.A.P., Leal, J.P.: PETCHA: a programming exercises teaching assistant. In: Proceedings of the 17th ACM Annual Conference on Innovation and Technology in Computer Science Education, pp. 192–197 (2012)
Tang, C.M., Yu, Y.T., Poon, C.K.: An approach towards automatic testing of student programs using token patterns. In: Proceedings of the Seventeenth International Conference on Computers in Education (ICCE), pp. 188–190 (2009)
Tang, C.M., Yu, Y.T., Poon, C.K.: An experimental prototype for automatically testing student programs using token patterns. In: Proceedings of the International Conference on Computer Supported Education (CSEDU), pp. 144–149 (2010)
Wasik, S., Antczak, M., Badura, J., Laskowski, A., Sternal, T.: A survey on online judge systems and their applications. ACM Comput. Surv. (CSUR) **51**(1), 1–34 (2018)

Automatic Question Generation System for English Reading Comprehension

Yin-Chun Fung[1(✉)], Jason Chun-Wai Kwok[1], Lap-Kei Lee[1], Kwok Tai Chui[1], and Leong Hou U[2]

[1] School of Science and Technology, The Open University of Hong Kong, Ho Man Tin, Kowloon, Hong Kong SAR, China
{ycfung,g1203482}@study.ouhk.edu.hk, {lklee,jktchui}@ouhk.edu.hk
[2] SKL of Internet of Things for Smart City, Department of Computer and Information Science, University of Macau, Macao SAR, China
ryanlhu@um.edu.mo

Abstract. This paper presents a web-based automatic question generation (AQG) system to generate reading comprehension questions and multiple-choice (MC) questions on grammar from a given English text. Such system saves teachers' time on setting questions and facilitates students and their parents to prepare self-learning exercises. Our web-based system can automatically generate Wh-questions (i.e., what, who, when, where, why, and how) and MC grammar questions of selected sentences. Wh-questions can also be generated from user-specified answer phrases. The generation of Wh-questions exploits the pre-trained natural language understanding model, Text-To-Text Transfer Transformer (T5), and an adapted version of the SQuAD 2.0 machine reading comprehension dataset. The generation of MC questions involves identifying regular verbs in a text and using the verb's lexemes as the answer choices. Our system takes an average time of about 1 s to generate a Wh-question and it generates a MC question almost instantly. User evaluation indicated that our system is easy-to-use and satisfactory in usefulness, usability, and quality, revealing the effectiveness of our system for teachers and parents.

Keywords: Automatic question generation · English reading comprehension · Wh-questions · Multiple-choice questions · Natural language processing

1 Introduction

Language learning is compulsory in most schools. According to Cartwright (2002), reading is a cognitive demanding task. Reading comprehension can facilitate the development of cognitive skills and the learning of new vocabulary (Nagy 1988). Hence, reading comprehension is commonly adopted in the language learning process. Teachers give reading comprehension as assessment in schools. There are numbers of online platforms for students to practice reading comprehension nowadays. Some parents would like to set up exercises for their children to train their reading skills. There are

L.-K. Lee et al. (Eds.): ICTE 2020, CCIS 1302, pp. 136–146, 2020.
https://doi.org/10.1007/978-981-33-4594-2_12

a lot of articles on the Internet that can be used as reading comprehension material. However, setting up questions is a time-consuming and labour-intensive task (Mitkov et al. 2006). Recently, some researchers focused on automatic question generation (AQG) (see Kurdi et al. (2020) and the references therein). Yet, it is rare to find an existing system that is easy and free for non-technical users to generate reading comprehension questions in a fast and massive way. In addition, English has long been one of the most important language, supported by many research articles (Bury and Oka 2017; Qi 2016). Thus, we have decided to implement a web-based system to provide an easy-to-use platform for teachers and parents of junior students to generate English reading comprehension exercises.

1.1 Existing Approach on AQG

There are numbers of algorithms on automatic question generation (AQG). We can generally classify them into three categories: template-based, syntax-based, and semantics-based (Kurdi et al. 2020; Yao et al. 2012):

- **Template-based.** The first step of template-based approach is to define templates consisting of some fixed text, such as 'What is X' and 'When did X begin'. The main purpose of this category of AQG algorithm is to find suitable keywords in a text to substitute 'X' into the template (Liu et al. 2018).
- **Syntax-based.** Syntax-based approach makes use of syntax structure of sentences in a text. Syntax rules and transformation are defined. If there are sentences that match the syntax rules, questions can be generated using transformation rules (Heilman and Smith 2009).
- **Semantics-based.** Semantic features are analysed in semantics-based approach. By recognizing the semantic meaning between phrases, questions are generated. (Flor and Riordan 2018).

Recently, more studies are using sequence to sequence (seq2seq) encoder-decoder model with attention mechanism on AQG (Du et al. 2017; Zhou et al. 2018; Yuan et al. 2017; Zhao et al. 2018; Hosking and Riedel 2019). Seq2seq, introduced by Google (Sutskever et al. 2014), means that the input and output are both sequences. Encoder-decoder is an architecture that combines the encoder and decoder network. The encoder is a network that turns the input into a vector containing the information/features of the input. The decoder is the opposite of encoder, which is a network that turns the vector into an output item. They have usually been employed together as encoder-decoder architecture. Attention mechanism (Bahdanau et al. 2015) is often used with encoder-decoder to improve the performance. It allows the model to pay attention to the most important part of the text instead of all the text.

Transformer-based models (Vaswani et al. 2017) are also popular in AQG nowadays. It is an encoder-decoder architecture with a multi-head self-attention mechanism. Nearly all state-of-the-art models are based on the transformer. Bidirectional encoder representations from transformers (BERT) (Devlin et al. 2019) is the most popular model in natural language understanding. It consists of multi-layers of transformer encoders. GPT-2 (Radford et al. 2019) is the most powerful language model for text generation. It is based on the transformer decoder. These examples show how powerful and inspirational the transformer is.

Unlike the typical neural network model that requires to be trained from scratch, most of the transformer-based models are general-purpose language models that make use of transfer learning. These models are following a pre-training and fine-tuning process. In the pre-training stage, the model is trained with an extremely large corpus, so that the model is able to have a better representation of the language. This step usually requires a lot of computational resources. It can cost thousands of dollars per training. Therefore, researchers and organizations who proposed the models usually release the pre-trained models. Users can download the pre-trained models and perform fine-tuning for different downstream tasks with much less computational cost, while enjoying the performance gain from pre-training. It is worth noting that the fine-tuning process is equivalent to the tuning of hyperparameters. It is the training process that trains the model with the data for a specific task.

1.2 Existing System for AQG

Quillionz is an artificial intelligence (AI)-powered question generator. It provides both free and paid services. Free service provides only True-False questions, multiple choices questions and fill-in-the-blank questions, while Wh-questions and interpretive questions generation are paid service (Fig. 1). Users need to input a text between 300 and 3000 words and select the domain of the text. The system will initialize some keywords for the question generation process. Users can choose to include or exclude some of the keywords. It then requires users to review the content, which includes solving lengthy sentences, resolving pronouns, and modifying some subjective or incomplete sentences. This process aims to improve the quality of generated questions. Finally, questions will be generated.

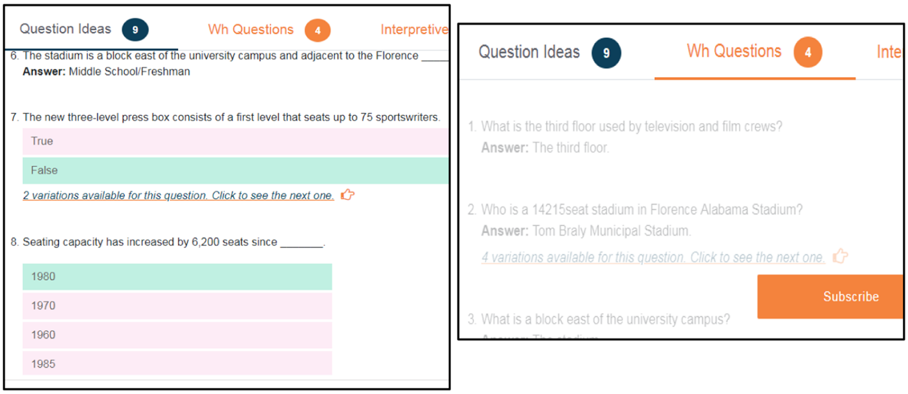


Fig. 1. Quillionz generates True-False, multiple choices and fill-in-the-blank questions for free (left). Wh- and interpretive questions generation are paid service (right).

The limitations of Quillionz are the minimum word count and the control of the content. The system rejects a text with less than 300 words which is not flexible for generating reading comprehension questions for junior students. If the text is not in the listed domain of the system, the generated questions may not be of good quality. Moreover, the system requires human effort to modify the text to fit the system in order to generate good quality questions. It is not user-friendly when the users are asked to "rewrite" the text but they do not think the text is that poorly written.

2 Our Web Application for AQG

Our AQG system consists of two major QG components: Wh-question generation and grammar question generation. When using the system, users need to input a text. The text will be processed in two ways to produce Wh-questions and multiple choice questions on grammar, respectively. Figure 2 shows the overall design of our system.

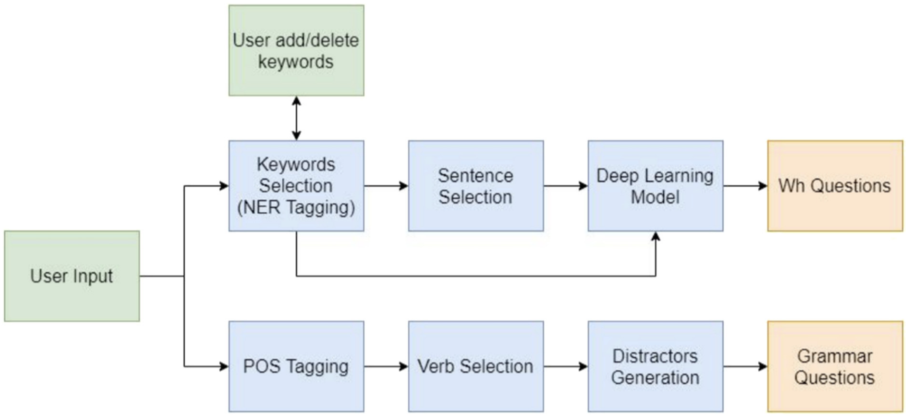


Fig. 2. The overall design of our web-based AQG system.

2.1 Grammar Question Generator

Our grammar questions are multiple choice questions on tenses. The first step is to identify which are the verbs in the text. We used a part-of-speech (POS) tagger (De Smedt and Daelemans 2012) to identify the part of speech of all individual words in the text. We choose all the verbs and produce its other lexemes as the distractors (i.e., wrong answers) of the question. If the number of lexemes is less than three for a particular verb, this verb will be ignored. There is no pre-training in this part of the generation process.

2.2 Wh-Question Generator

We have used a deep learning approach to generate Wh-questions. A pre-trained English model, Text-To-Text Transfer Transformer (T5), was adopted (Raffel et al. 2019) as the base model. Considering the processing power and the responding time of the system, we use “t5-small”, the smallest model of the T5 family, in our work. We fine-tune the T5 model using the benchmark SQuAD 2.0 dataset (Rajpurkar et al. 2018) to make it suitable for generating questions. Since the SQuAD 2.0 dataset is for question answering, we simply treat the text, question-answer pairs of the dataset as the inputs and outputs for fine-tuning our model to better fit the question generation task: the input is a source sentence with an answer phrase, whereas the output is a Wh-question.

When using our system, the input text will go through a named-entity tagging process to find the possible answer phrases. Then, the selected phrases, called keywords, will be passed into our fine-tuned model together with the source sentence. A Wh-question will then be produced as the output.

2.3 Wrapping the Two Generators

We used a web application to wrap the whole process. To generate questions, users will be required to input a text (Fig. 3). The system will select and display the possible answer phrases to the users. Users can choose new keywords and un-select any keyword according to their needs (Fig. 4).

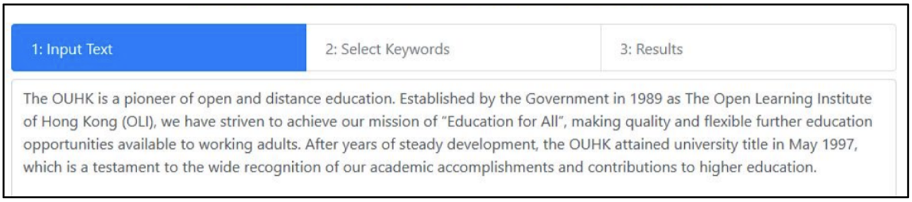


Fig. 3. First step of using our system: input a text.

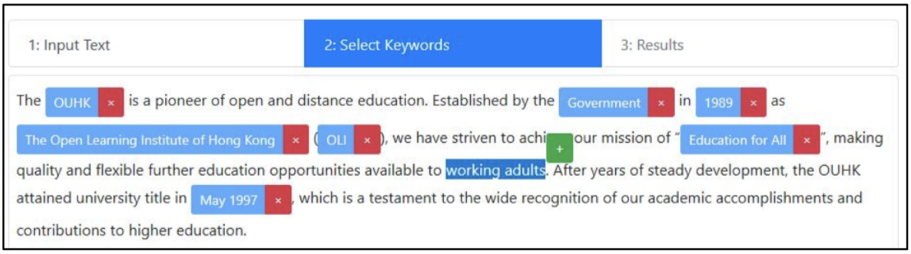


Fig. 4. Second step of using our system: select keywords.

When all settings are ready, questions will be generated automatically (Fig. 5), as introduced in Sects. 2.1 and 2.2. Generated Wh-questions are shown on the left in Fig. 5. Users can click on the “Source” button to see the original sentence in case they want to make sure the questions are set correctly, or would like to examine the answer (Fig. 6). The right of Fig. 5 shows the generated multiple choice questions on grammar. Users can shuffle the choices of the questions by clicking the “shuffle” button.

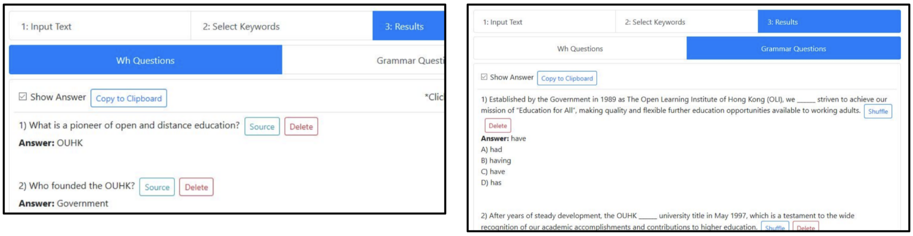


Fig. 5. Wh-questions (left) and Grammar multiple choice questions (right) are generated.

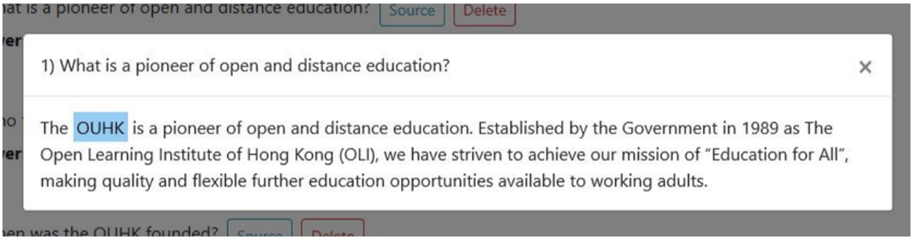


Fig. 6. Users can view the source of the generated question.

In both question pages (Fig. 5), a user would remove some questions directly using the "Delete" button next to those questions. All answers will be hidden by un-selecting the "Show Answer" option. The "Copy to Clipboard" button allows users to copy the questions to other file for further processing. Users can also print the web page directly.

3 Evaluation

3.1 Response Time

We conducted a response time test to evaluate the time needed to generate a question using GPU and CPU. There are 5 randomly chosen texts, the number of words and number of questions is shown in Table 1 below. We collected the time used to generate questions in these 5 texts, and the average time needed per question, and tried running the model on CPU (i7-4770) and GPU (RTX 2060).

The result shows that the average time cost per question using CPU is 1.3 s, and that of GPU is 1.07 s. It shows that GPU outperforms CPU by 18% in AQG.

Note that the time for the NER tagging and generating grammar questions is very short (much less than one second) and are therefore omitted from the response time test. To sum up, our web-based AQG system provides satisfactory response time on the question generation.

3.2 Quality of Generated Questions

We evaluated the capability of the system in generating different types of questions, including "What", "Who", "When", "Where", "Why", "How", and "How many" (Table 2).

Table 1. Result of response time test.

	No. of words	No. of questions	Time used (second)		Avg. time cost per question (second)	
			CPU	GPU	CPU	GPU
Text 1	312	26	35.1	28.7	1.35	1.10
Text 2	180	25	32.0	26.5	1.28	1.06
Text 3	340	18	23.1	19.0	1.28	1.06
Text 4	160	19	25.8	20.5	1.36	1.08
Text 5	117	23	28.7	24.6	1.25	1.07
Mean average time cost per question (second)					1.30	1.07

The results indicated that the system can generate different types of Wh-questions if the answers are selected properly. Our auto-selected answers are based on NER tagging, which can select the answers for generating “What”, “Who”, “When”, “Where”, and “How many” type of questions. Yet it is not suitable for automatically generating the “Why” and “How” type of questions. The generation of these types of questions highly relies on user involvement and thus the user may need to manually select the answers to ensure the quality of the questions.

3.3 User Survey

To evaluate the performance of our web-based AQG system, we invited 15 participants to test the system and complete a survey. The survey has 8 questions, where Questions 1 to 7 used a 5-point Likert scale (1: strongly disagree, 2: disagree, 3: neutral, 4: agree, 5: strongly agree) and are on the usefulness, usability, and look-and-feel of the system, and Question 8 is an open-ended question. Table 3 shows the result on Questions 1–7.

The first two questions are on the usefulness of our system. The result shows that most of the users are satisfied with the quality of the questions. However, we can see the satisfaction of grammar questions is higher than Wh-questions, which implies a room for improvement in the Wh-questions generation component.

Questions 3 and 4 are on the usability of our system. It reflects if the design of the system and user interface is logical. The result shows that a majority (over 80%) of users gave the highest rating 5 on this part.

Questions 5 and 6 are on the look-and-feel of the system, which is mostly related to the user interface. The result shows most of the users are satisfied with the user interface, but there is still a little room for improvement since there are still 33% of the users who did not select the highest score. We expect adding more instructions and functions to the user interface could be useful.

As indicated in the result of Question 7, all users were satisfied with our system.

Table 2. Result of question type test.

Type		
What	Content	Amazon is the most valuable brand in the world.
	Selected Keyword	Amazon
	Generated Question	What is the most valuable brand in the world?
Who	Content	Donald Trump is the current president of the United States
	Selected Keyword	Donald Trump
	Generated Question	Who is the current president of the United States?
When	Content	Donald Trump was born on June 14, 1946
	Selected Keyword	June 14, 1946
	Generated Question	When was Donald Trump born?
Where	Content	Donald Trump was born at the Jamaica Hospital in the borough of Queens, New York City
	Selected Keyword	Jamaica Hospital
	Generated Question	Where was Donald Trump born?
Why	Content	The increased solubility of O2 at lower temperatures (see Physical properties) has important implications for ocean life, as polar oceans support a much higher density of life due to their higher oxygen content
	Selected Keyword	higher oxygen content
	Generated Question	Why do polar oceans support a much higher density of life?
How	Content	The principal Treaties that form the European Union began with common rules for coal and steel, and then atomic energy, but more complete and formal institutions were established through the Treaty of Rome 1957 and the Maastricht Treaty 1992 (now: TFEU)
	Selected Keyword	with common rules for coal and steel
	Generated Question	How did the Treaties that form the European Union begin?
How many	Content	Only a few common complex biomolecules, such as squalene and the carotenes, contain no oxygen
	Selected Keyword	A few
	Generated Question	How many complex biomolecules contain no oxygen?

Question 8 is an open-ended question for users to comment on the system. It aims at collecting users' feedback to improve the limitations of the system. There are three main limitations: (i) The first one concerns that the auto-selected keywords are not good enough. The keywords are selected based on NER tagging. It is possible that the whole input text has no or very little named-entity. We also noticed that it is a challenging task to generate the "why" question if the answer is a named-entity. Therefore, it is one of

Table 3. Result of user survey on Questions 1 to 7.

Question		Percentage (%)				
		1	2	3	4	5
1	I am satisfied with the quality of Wh-questions	0	0	13.3	40	46.7
2	I am satisfied with the quality of grammar questions	0	0	0	13.3	86.7
3	I think the system is easy to use	0	0	0	13.3	86.7
4	I think the system is easy to learn	0	0	0	20	80
5	The system works the way I expected	0	0	0	33.3	66.7
6	I like the user interface of the system	0	0	0	33.3	66.7
7	Overall, I am satisfied with the system	0	0	0	46.7	53.3

the key issues to be solved in the future; (ii) the second one concerns the long loading time. The major reason for the long loading time is that the model we used is large. Powerful hardware or optimizing the program could help with this problem; and (iii) the last one concerns the quality of the generated questions. The quality of the questions mostly relies on the model. There is a trade-off of performance and response time.

4 Conclusion and Future Work

An easy-to-use and web-based English reading comprehension question generation system has been built. It can generate Wh-questions and multiple choice grammar questions on tenses. Analysis revealed that the question generation takes has a satisfactory response time (about one second per question). Survey was conducted and participants were satisfied with the current system. Though, there is room for improvement in future work. Future research directions are suggested in three perspectives.

First of all, considering the quality of generated questions. The auto-selected keywords need to include those other than named entities such that the "why" type questions can be generated without human involvement. Moreover, more questions types, like part of speech and open-ended questions, can be included in the future. If the difficulty of the questions can be increased, the system will be useful to senior students and their teachers and parents.

Second, the response time should be reduced. When processing a long text, the response time is not impressive. Rather than using a pre-trained model, we can train a model just for AQG from scratch. In future, we may also use other new models that are much powerful but with smaller size than “t5-small”.

Third, we can mimic similar approaches for different languages such as Chinese. Reading comprehension is also playing an important role in other languages. If different languages can be supported, more people can enjoy the convenience of AQG.

Looking to the future, we believe that AQG on reading comprehension can be greatly helpful to teachers, parents, and students. Teachers can save their time on preparing teaching materials. Parents can prepare more exercise for their children and the children as students can learn more from the exercises. We look forward to the well development on AQG.

References

Bahdanau, D., Cho, K., Bengio, Y.: Neural machine translation by jointly learning to align and translate. In: 3rd International Conference on Learning Representations, ICLR 2015, January 2015

Bury, J., Oka, T.: Undergraduate students’ perceptions of the importance of English in the tourism and hospitality industry. J. Teach. Travel Tour. **17**(3), 173–188 (2017)

Cartwright, K.B.: Cognitive development and reading: the relation of reading-specific multiple classification skill to reading comprehension in elementary school children. J. Educ. Psychol. **94**(1), 56 (2002)

De Smedt, T., Daelemans, W.: Pattern for Python. J. Mach. Learn. Res. **13**, 2031–2035 (2012)

Devlin, J., Chang, M.W., Lee, K., Toutanova, K.: BERT: pre-training of deep bidirectional transformers for language understanding. In: NAACL-HLT (1), January 2019

Du, X., Shao, J., Cardie, C.: Learning to ask: neural question generation for reading comprehension. In: Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pp. 1342–1352, July 2017

Flor, M., Riordan, B.: A semantic role-based approach to open-domain automatic question generation. In: Proceedings of the Thirteenth Workshop on Innovative Use of NLP for Building Educational Applications, pp. 254–263, June 2018

Heilman, M., Smith, N.A.: Question generation via overgenerating transformations and ranking (No. CMU-LTI-09–013). Carnegie-Mellon Univ Pittsburgh pa language technologies insT (2009)

Hosking, T., Riedel, S.: Evaluating rewards for question generation models. In: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), pp. 2278–2283, June 2019

Kurdi, G., Leo, J., Parsia, B., Sattler, U., Al-Emari, S.: A systematic review of automatic question generation for educational purposes. Int. J. Artif. Intell. Educ. **30**(1), 121–204 (2020)

Qi, G.Y.: The importance of English in primary school education in China: perceptions of students. Multiling. Educ. **6**(1), 1–18 (2016). https://doi.org/10.1186/s13616-016-0026-0

Liu, T., Wei, B., Chang, B., Sui, Z.: Large-scale simple question generation by template-based seq2seq learning. In: Huang, X., Jiang, J., Zhao, D., Feng, Y., Hong, Yu. (eds.) NLPCC 2017. LNCS (LNAI), vol. 10619, pp. 75–87. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-73618-1_7

Mitkov, R., Le An, H., Karamanis, N.: A computer-aided environment for generating multiple-choice test items. Nat. Lang. Eng. **12**(2), 177 (2006)

Nagy, W.E.: Teaching vocabulary to improve reading comprehension. National Council of Teachers of English, Urbana, IL.; ERIC Clearinghouse on Reading and Communication Skills, Urbana, IL.; International Reading Association, Newark, DE (1988)

Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., Sutskever, I.: Language models are unsupervised multitask learners. OpenAI Blog **1**(8), 9 (2019)

Raffel, C., et al.: Exploring the limits of transfer learning with a unified text-to-text transformer (2019). arXiv preprint arXiv:1910.10683

Rajpurkar, P., Jia, R., Liang, P.: Know what you don't know: unanswerable questions for SQuAD. In: Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers), pp. 784–789, July 2018

Sutskever, I., Vinyals, O., Le, Q.V.: Sequence to sequence learning with neural networks. In: Advances in Neural Information Processing Systems, pp. 3104–3112 (2014)

Vaswani, A., et al.: Attention is all you need. In: Advances in Neural Information Processing Systems, pp. 5998–6008 (2017)

Yao, X., Bouma, G., Zhang, Y.: Semantics-based question generation and implementation. Dialogue Discourse **3**(2), 11–42 (2012)

Yuan, X., et al.:Machine comprehension by text-to-text neural question generation. In: Proceedings of the 2nd Workshop on Representation Learning for NLP, pp. 15–25, August 2017

Zhao, Y., Ni, X., Ding, Y., Ke, Q.: Paragraph-level neural question generation with maxout pointer and gated self-attention networks. In: Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, pp. 3901–3910, October 2018

Zhou, Q., Yang, N., Wei, F., Tan, C., Bao, H., Zhou, M.: Neural question generation from text: a preliminary study. In: Huang, X., Jiang, J., Zhao, D., Feng, Y., Hong, Yu. (eds.) NLPCC 2017. LNCS (LNAI), vol. 10619, pp. 662–671. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-73618-1_56

Learning Environment

Social Media Acceptance Scheme

Miloslava Cerna(✉)

Faculty of Informatics and Management Hradec Kralove, University of Hradec Kralove, Hradec Kralove, Czech Republic
miloslava.cerna@uhk.cz

Abstract. The longitudinal study on acceptance of social media in the university setting brings a decade experience on how students perceive, use and value selected social media for study purposes. Based on literature review on utilization of social media in education 'Social media acceptance scheme' was designed and repeatedly applied with slight relevant adaptations reflecting researchers targets of conducted national and international university projects within the discussed time span of ten years. Research targets vary, however awareness of social media platforms represents a kind of continuum being of key importance when considering and creating a digital strategy in the process of education at various levels of its system. Based on the findings gained from 'Social media acceptance scheme', the concept of blended learning has been adapted accordingly; study materials in e-courses were reorganized as well as innovative approaches in face2face classes have been made in teaching/learning Professional English language and Czech language for foreigners subjects at the University of Hradec Kralove. The latest data on utilization of social media by Czech and foreign students bring an international dimension where comparison of findings is visualized and discussed.

Highlight. 10 years' experience on the local scene within the global frame with an intercultural overlap on acceptance, utilization and potential of social media in university setting.

Keywords: Social media · Research · Student engagement · Student satisfaction · Longitudinal study

1 Introduction

The worldwide accessibility to the Internet is one of the defining phenomena of the present times in reshaping the world. Clement (2020) calls Social media the lovechild of the World Wide Web. Social media with their wide range of forms like blogs, forums, and business networks, sharing platforms, social gaming, microblogs, chat applications or social networks develop, adapt to new requirements, stagnate or leave the scene. Due to their influential and defining role in the private and professional lives, social media have been in the centre of research interest since the birth of second generation of WWW so called Web 2.0. Selected relevant topics of social media research follow as an illustration of the issue scope and breadth:

L.-K. Lee et al. (Eds.): ICTE 2020, CCIS 1302, pp. 149–163, 2020.
https://doi.org/10.1007/978-981-33-4594-2_13

- the challenges of social media use and research (Osch and Coursaris 2015; Weller 2015, 2016),
- intra-generational variance (Bolton et al. 2013),
- social media marketing (Tuten and Solomon 2017),
- social media engagement (Schivinski et al. 2016),
- and many other areas including the area of education, which is in focus of this paper (Greenhow and Lewin 2016; Selwyn 2017), see more in the Relevant studies subchapter.

The longitudinal study on the acceptance of social media in the university setting brings a decade experience on how students perceive, use and value selected social media for study purposes.

The structure of the paper follows the standard pattern: State of Art consisting of subchapters on social media statistics, categorization and relevant studies within the methodological frame of the research, Design of the Social Media Acceptance Scheme, Findings, and Conclusion.

2 State of Art

The chapter deals with the role of social media and literature sources. *Statistics on selected statistical websites is used as one of the approaches to the research on utilization of social media and forms the first opening part of the State of Art chapter*.

The direction goes from global to local; it is shifted from the wide global frame to the applied research on the local scene.

2.1 Social Media on the Global, European and Local Scene in the Statistical Numbers

This subchapter brings a brief statistical overview on social media users and most popular social networks. Proven relevant websites providing *global data* and their processing on the researched issue are: Pew Research Center (2020) with a research area Internet & Technology; Internet World Stats (2020), IWS for short, website for international Internet usage statistics, population Statistics, Social Media Stats and Internet Market Research Data, for nearly 250 countries; or Statista (2020) with its section Social media - Statistics & Facts.

As for the *local scene*, data were taken from two key sources: Eurostat (2016) and the Official Statistical Office of the Czech Republic (Czech Statistical Office 2020).

Statistical numbers show the enormous deal of social media users out of the total population. The role that social media play in their life can be seen in time people spend on the Internet from their waking time.

Kemp (2019) in the Global Digital 2019 report reveals that currently, the total population reaches 7.7 billion people; more than 5 billion are mobile users, 4. 4 billion are internet users and 3.5 billion are social media users. Social media user spends astonishing 2 h and 15 min on social platforms daily, which represents about one-third of their time on the Internet, and when look at that from another perspective this time represents one-seventh of their waking lives. However, *there are discrepancies in users' behaviour*. The numbers give the global view. In reality, what do they illustrate, what is the predictive value, to what extent can researchers built upon them?

According to Statista (2019), most popular social networks worldwide in October 2019 ranked by number of active users (in millions) follow: *Facebook* (2, 414), *Youtube* (2000), *WhatsApp* (1,600), Facebook Messenger (1,300), WeChat (1,133), *Instagram* (1,000), QQ (808), QZone (554), Douyin/Tik Tok (500), Sina Weibo (486), Reddit (330), *Twitter* (330), Douban (320), Snapchat (314), *LinkedIn* (310), Pinterest (300).

Photo-sharing is widely popular among students of our Faculty of Informatics and Management; *popularity of photo sharing is astonishing, to our surprise it is popular with foreign students in utilization of this tool even for study purposes, as will be discussed later in the chapter Findings.*

Another note also refers to intercultural dimension of social media. As for social networks, one of crucial characteristics is that they are in multiple languages, which enables its borderless spread; it is an enriching topic for raising discussion on comparison of these networks among students from various cultural backgrounds (Černá and Lin 2020). Another feature worth highlighting is the role of some social network in the local context. The great deal of social networks with more than 100 million users originated in the United States, but European services like VK (VKontakte, 'InContact') or Chinese social networks Qzone and Renren have also gained strong position on their local scene (Statista 2019). Behavior and habits of social media users in Europe look often similar as in other part of the world, which fosters the idea of global phenomenon. There are differences in social media preferences, e.g., Facebook is not so popular in Russia, VK holds its leading position with over 400 million users in Russia and parts of Europe.

Awareness of social media platforms is of key importance when considering and creating a digital strategy within various markets as well as in the educational sphere, which is in focus of this paper.

Lorenc (2018) in his report on utilization of social media channels in the Czech Republic gives statistics and demographic in brief. In the Czech Republic with 10.65 million inhabitants, following platforms dominate the social media scene. *Facebook Network* with its several platforms, e.g., Messenger with nearly 6 million users is the main player. It is followed by *YouTube* with 5 600 000 active users per month. Another big player is social media platform *Instagram* from Facebook family with 2.2 million users. Then *LinkedIn* comes with 1 600 000 career oriented users. Twitter's audience is even smaller consisting of 600 000 users.

The use of social networks in the countries of the European Union varies. On average, there are 55.8% social media users in the EU. The Czech Republic belongs to the 'precise' average with its 55.6% of users. The reached level of education gets reflected in the statistical overview: *75% university graduates use social media*, 55% secondary school graduates is followed by 41% of people who accomplished the basic level of education (Czech Statistical Office 2020).

2.2 Categorization of Social Media

There have been many concepts of categorization of tools brought by Web 2.0, which have had to change and adapt, as the development of technology does not stagnate; individual tools gain new functionalities and get more and more interconnected. Social media are standardly divided into two main sections: social networks like Facebook, Twitter, LinkedIn and media sharing sites, e.g., YouTube or Instagram. *However, there is a current trend where the lines between media sharing networks and social networks are blurring.* Many multimedia services, e.g., adding live video or augmented reality, are functionalities that are used also on platforms of Facebook, Google + or LinkedIn. Grouping social networks according to subject matter or the functional capabilities has become old fashioned. This is finding which was also reached during the follow – up discussions with students at the qualitative stage of the presented research on social media. Foreman (2017) shifts categorizing networks from their division based on technology features to categorizing networks based on the aim of people use.

2.3 Relevant Studies Within the Methodological Frame of the Research

This longitudinal study monitors and investigates students' acceptance, utilization and potential of social media in the university setting.

Studies presented in this subchapter deal with key characteristics of Web2.0, which play a distinguished role in the process of education. The studies roughly cover the time span corresponding to the time span of the longitudinal study. There are studies from which researchers took inspiration for the design of their research, especially the survey as a main research tool e.g., Selwyn (2007b) and Usluel and Mazman (2009). Following literature sources help to illustrate the development and trends in the research of social media in education.

Terminology relating to the interactive Web 2.0 technologies that appeared after static Web 1.0 might seem confusing. In older studies, terms like Web 2.0 technologies, Web2.0 applications, Web2.0 social software applications, Web 2.0 services dominate. *Currently terms like social platforms or social media prevail.* Our research can be an example of this shift; we started with the term social software applications and moved to social media.

With the emergence of 'Web 2.0' technologies, there was a great deal of enthusiasm and expectations about new possibilities in enhancing learning among educationalists. Selwyn (2007b) with his critical review on research literature on Web2.0 applications was inspiring for our research with his sober approach to actual practical utilization of these applications in facilitating new forms of creation, collaboration and consumption in learning. *Some of his questions concluding the paper have been in the foci of researchers up to now, e.g. "to what extent do learners expect/desire to use 'informal' forms of ICT use in the formal educational settings such as the school? Which ICTs do learners see as being most motivating, engaging and personalisable? Conversely, which ICTs do learners see as unsuitable for the classroom and why? What unintended consequences and/or risks do learners see as arising from importing 'new' informal modes of ICT use into the classroom setting?"* (Selwyn 2007b). Ten years later Selwyn (2017) published a comprehensive book on practices, processes and structures behind the rapidly increasing use of technologies in education, with an emphasis on the implications of digital technologies for individuals and institutions.

Usluel and Mazman (2009) discussed in their study new possibilities of *interaction*, which Web2.0 supports. They highlighted the *importance of individuals*' *acceptance and usage of technology* as the most effective ways of obtaining effective learning outcomes in distance education within the frames of various models and theories about diffusion, acceptance and adoption of an innovation. They suggested that Unified Theory of Acceptance and Use of Technology (Venkatesh et al. 2003) would bring out more in depth and comprehensive approach.

Anderson presents his idea of Web 2.0 on the iceberg model with social web on the top of the iceberg. He provides readers with implications of computing, and he focuses on repeatedly mentioned key characteristics of Web 2.0, e.g., *user-generated content, ways of participation or openness* (Anderson 2012). What we call social software applications he calls Web 2.0 services; out of these services, he *discusses blogs, wikis, social networks, media sharing sites, social bookmarking and microblogging*. His selection slightly differs from ours, as will be shown in the core chapter 'Design of the Social Media Acceptance Scheme'. He looks at how each of selected services is used, how it was developed and what technology is involved finally he adds important research themes and findings from the literature (Anderson 2012).

*We focus on the issue from students' perspective: what they know, what they use, how they are satisfied with the services and what potential students see in them for study purposes, which services they find appropriate for the use and management of study materials, possibilities of communication and testing knowledge for communication, revision, e.g., Černá and Poulová (*2013*).*

Key characteristics of Web2.0 seem incredibly promising for teaching/learning purposes. However, there is also the other side of the coin. WWW web has become a place of user participation and user interaction, with enormously increasing amount of material. The amount of user-generated content on the web is rising and comes from various sources. Here arises the issue of handling the available information appropriately. To obtain the right content at the right time and to reduce information overload effects is becoming a greater challenge for Web users. Students are offered so much study material which in many cases leads to the Paradox of Choice or the Decision Paralysis (Schwartz and Johnson 2013). We solved this issue of necessity to sort out study material in the Learning Management System (LMS) Blackboard, which is one of the analysed social platforms in this paper, via creating a modified recommender system model (Černá and Borkovcová 2018; Černá and Lin 2020) solved this problem, as well. In their paper on users' acceptance of Web2.0 services in academia, they highlighted the concept of information literacy competency.

When we return to the positive aspects of social media, it is evident that possibility of *content development* belongs to the best ways of engaging students into the process of education. Steyn et al. (2017) believe that students can better grasp the content and increase their learning via utilization of virtual platform and working out own material there. Researchers consider working in Blackboard to be natural, because their students are millennials, the generation growing up with the Internet. *The aspect of engagement* into the learning in social media has been part of our research since its beginning. However, findings on active involvement (Černá and Borkovcová 2019) correspond to less enthusiastic findings of (Greenhow and Lewin 2016) in their study on bridging formal and informal learning where belief in social media *potential* is discussed. They proposed a model theorizing social media as a space for learning with varying attributes of formality and informality. They are also more sceptical about active engagement as *they claim that majority of young people act as consumers rather than full participants* (Greenhow and Lewin 2016). Our findings also revealed that active cooperation or even collaboration is closer to wish than reality (Černá 2017; Černá and Borkovcová 2018).

As for academic sphere, according to (Weller 2015) some social media platforms are more prominent in academic publications than others are. She has long been involved in interdisciplinary nature of social media research. She claims that Facebook, Twitter, Wikipedia or YouTube due to their popularity are more analysed, especially user motivation and usage scenarios, then, e.g., Instagram (Weller 2015). Except for Twitter, our findings prove that Facebook, Wikipedia and YouTube belong to most prominent by students in our research samples. Weller (2016) focuses on social media platforms from user' perspective and finds out that some features of popular social media platforms like hashtags are currently understudied. She discusses the role of social medial key features in her research on understanding users' behaviour. Our research is also run from (users') students' perspective. Cheung (2016, 2017) utilizes features of Web 2.0 in teaching/learning process in his long-term research. Two of papers have been selected studies as they fit the concept of here presented longitudinal study. He was involved in the design and implementation of a complex project Open textbooks on on-line platforms (Cheung 2016). The other study deals with students' perspective; students' perception on usefulness of open educational resources is analysed (Cheung 2017).

3 Design of the Social Media Acceptance Scheme

The chapter is opened with the set of social media accompanied with a brief justification of their selection into the research. Then the individual stages of the design of the Social Media Acceptance Scheme are described. Examples of published outcomes of the research on utilization of social media, where readers can gain further information, conclude the chapter.

3.1 Categorization of Social Software Applications Applied in the Research

We have to admit, that Foreman's categorization (2017) only slightly differs from the categorization, which we designed 10 years ago. Selection of social software applications for our research purposes stemmed from studies discussing experience in utilization of these Web 2.0 tools especially in university setting. *The inspiring studies were chosen due to their balanced approach to the issue.* There could be seen not only pure optimism in technical innovations possessing motivating drive in young people to participate and cooperate in new channels in virtual space but also pitfalls and weakness, clarifying ways of implementing this phenomenon into the process of education were presented (Schroeder et al. 2010; Usluel and Mazman 2009; Weller et al. 2010). 9 kinds of social applications with defined key missions were placed into the *questionnaire* and analysed in the first run of the survey (Černá et al. 2011a, b).

- First monitored social applications were *Social nets* like Facebook and Twitter as social utilities connecting people with main mission that is communication, three years later Google+ was added to these nets (Černá 2014).
- Then *YouTube* as an application enabling sharing predominantly music recordings as well as presentations and instructions from various areas was selected. With audio and visual characteristics, this social application was considered a strong player on the virtual educational scene from the very beginning. Its dominance is strong and firm.
- *Wikis*, which can serve as platforms for knowledge integration and cooperation developing common knowledge, still keep their positions among students and enthusiastic educationalists. The fact is that there is a great deal of academics who consider wikis not sufficient or appropriate for study purposes due to a considerable high amount of unverified data and frequent simplification of explained issues.
- *Skype* and *ICQ* were selected as exclusively social communication application for chatting. ICQ has already left the active social application scene, however we left it in the questionnaire as an old-fashioned app fitting discussions on development of media from archaic to new platforms like *ZOOM*.
- *Blog* as a form of online reflective diaries was placed into the questionnaire because this social application had been found quite promising for study purposes in related literature and fully fitting our research (Schroeder et al. 2010; Selwyn 2007a; Weller et al. 2010). However, findings from our research do not correspond to findings in related literature, see more, e.g., (Černá 2014).
- *Social-bookmarking* fitting for storing and sharing web-links was found by academics highly beneficial for study purposes but findings from our research were even worse than in case of Blogs.
- *Sharing photos* (on-line photo albums) applications have undergone crucial changes from wide utilization in 2010 to the shift to new applications which are often part of dominant social networks.
- In the first runs of the longitudinal study there was a relatively large space given to *on-line computer games*, namely to World of Warcraft and Second-life, then there was also an open question if students knew, used and seen potential for study purposes in any other computer games. As the findings were marginal, original three questions in the questionnaire were condensed into one under general name on-line computer games.
- The last questions in the questionnaire dealt with Learning management systems (LMS) like WebCT (now Blackboard) and Moodle and other potential LMS. The individual Institutions often host these robust systems where they form an integral part of the overall course management.

3.2 Design of the Social Media Acceptance Scheme

The set of applications that form the core of the research has been introduced. Now a detailed description of individual sections of the Social Media Acceptance Scheme follows. This scheme was elaborated chronologically within the frame construct: *Awareness – Utilization – Satisfaction – Potential*.

Quantitative data are gained from the Questionnaire filled by students where defined areas are examined. Firstly, there is *Awareness* – students mark whether they know the application. If students know the application, it doesn't guarantee that they use it. So the next step is *Utilization* – when students mark that they use the app, they mark *Frequency* - how often they use it. The following step is *Purpose* of the usage of the app. Our research is designed for tertiary education that is why students are asked to mark if they use the individual apps for the purposes of *further education*, for the *university study purposes* or for study purposes in both areas. In case they mark University study purposes, a following important stage opens which deals with the appropriateness of the apps for specific parts of the learning process: *Reservoir of study materials, Communication and Revision.* Students select the level of *Satisfaction* with the apps from the 1–5 scale. Based on the level of satisfaction the *Potential* of individual applications for study purposes perceived by students might be identified. The interest in the potential of social media discussed in the follow-up discussions led the researchers to adding *Potential* as a separate category into the questionnaire.

Visualisation of the Social Media Acceptance Scheme can be seen in Fig. 1.

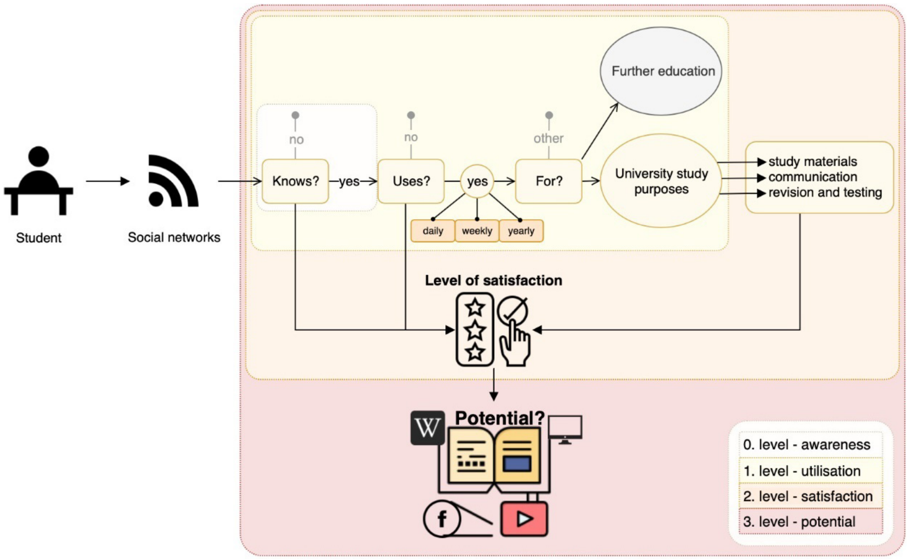


Fig. 1. Social media acceptance scheme

The *follow-up discussion* with respondents is beneficial in collecting data and opinions; it enriches investigation into qualitative space. The topic is so natural to students that they easily enter the discussion and are ready to share their experience even if there is a slight language barrier. Showing them findings from other cultural backgrounds rises their involvement. Comparison works as a proved motivating factor for engagement into a talk (Černá and Lin 2020).

Examples of published outcomes of the research on utilization of social media, where readers can gain further information follow. Social media scheme has been *applied repeatedly*, it has been adapted to fit the goals of researchers. One of the first outputs discussing the *whole scene of researched* media was 'Role of social media in academic setting awareness, utilization and willingness' (Černá and Poulová 2013). Examples with specified focus follow: *Awareness and Potential* (Černá et al. 2011), (Černá and Svobodová 2018), *Trends in utilization of social media in further education and in university setting* (Černá 2014; Černá 2017), *Focus on one application,* e.g., Facebook (Černá et al. 2017), *Cooperation and Involvement* (Černá 2015), *(Černá and Borkovcová 2019), International overlap* – comparison of findings between research samples consisting of students from the Faculty of Informatics and Management, University of hradec Králové and Polish students from Wroclaw School of Banking (Černá and Poulová 2013), students from Karaganda State Technical University, Kazakhstan (Černá et al. 2014) or students from National Taiwan University of Science and Technology (Černá and Lin 2020).

4 Findings

Findings in this chapter are opened with a general overview of utilization of social media in a time line of one decade. Data were taken from two studies (Černá 2014; Černá et al. 2011a, b) and a survey conduced this year with research samples consisting of full-time students from the Faculty of Informatics and Management.

Findings visualizing the one decade time-line refer to two areas, which are general one and a specialised one. A general view can be seen in Fig. 2 showing Awareness of social media. The special focus deals with four social media fitting study purposes most, which is Fig. 3 Frequency of used media.

The other part of findings is devoted to latest survey conducted at the end of winter semester 2019/2020 with Czech and Foreign students attending classes Professional English language and Czech language for foreigners subjects at the University of Hradec Kralove. The latest data on utilization of social media by Czech and foreign students bring an international dimension where findings on Awareness, Satisfaction and Potential are compared, see Fig. 4, 5 and 6. It is necessary to keep in mind that there are *limitations* in this study connected with the decreasing number of respondents in research samples (257 (2010) 71 (2014) 38 (2020) which means that no generalization of findings can be made. However, *findings are authentic and represent actual state on a local scene.*

Dominance of Facebook, YouTube, Skype and Games can be seen at first glance in the Fig. 2. Interesting thing is that Bookmarking as a highly promoted social media application by academics in their studies hasn't been found by our students. Similarly, Blog is also 'prominent' social media used in education, but not so much by respondents in our research samples. According to the findings from the follow-up discussions, students travelling abroad within Erasmus study program appreciate Blogs as authentic sources

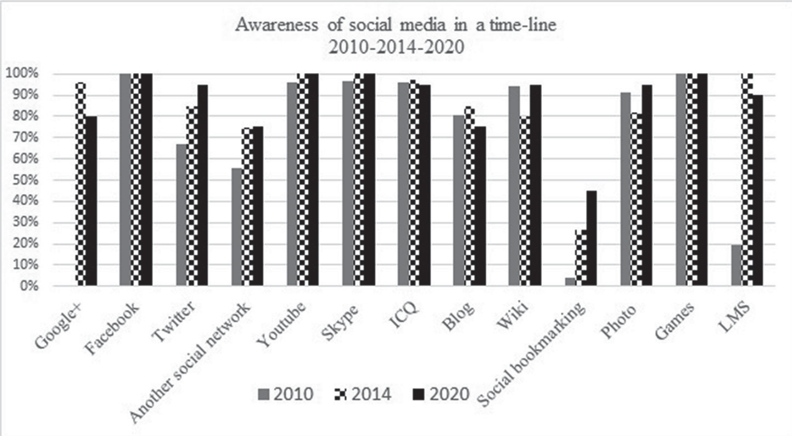


Fig. 2. Awareness of social media in a time line

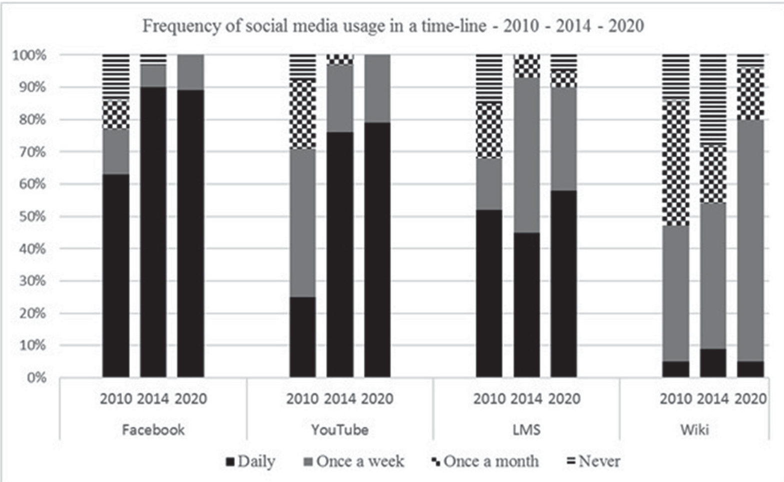


Fig. 3. Frequency of social media use in a time line

with authentic practical information. Skype is used for teaching/learning purposes in case of a student's absence due to a long-term illness and especially in communication with students who are abroad.

This year, due to a coronavirus pandemic, we switched to the school promoted Microsoft Teams platform. All Czech students know the almost extinct *ICQ* application, but no one from this year's group uses it anymore. In spite of the fact that *Learning Management System* was in 2010 widely used, only one fifth of students marked that they knew it, in 2014 all students marked that they knew LMS.

Figure 3 visualizes *Frequency* of use of researched social media as an important category for potential implementation into the process of education both either formal or

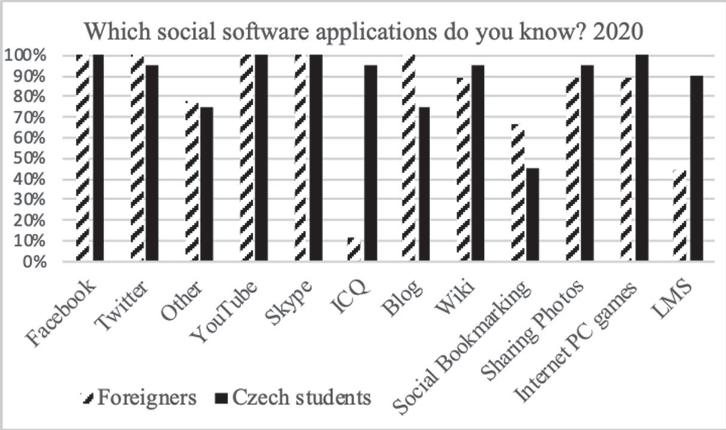


Fig. 4. Awareness of social media – comparison Czech a Foreign students

	G+	FB	Twitter	Instagr	Other	YouTube	Skype	ICQ	Blog	WA	Viber	Wiki	Bmark	Photos	Games	LMS
Czech	2,6	2,5	2,3		1,5	1,8	2,8	3,3	2,3			2,4	2,7	2,4	2,1	2,6
Foreign		2,7	2,6	1,8		1,1	2,8	5	3,7	1,4	5	1,9	2,6	1,6	2,8	3,5

Fig. 5. Satisfaction – comparison Czech a Foreign students view

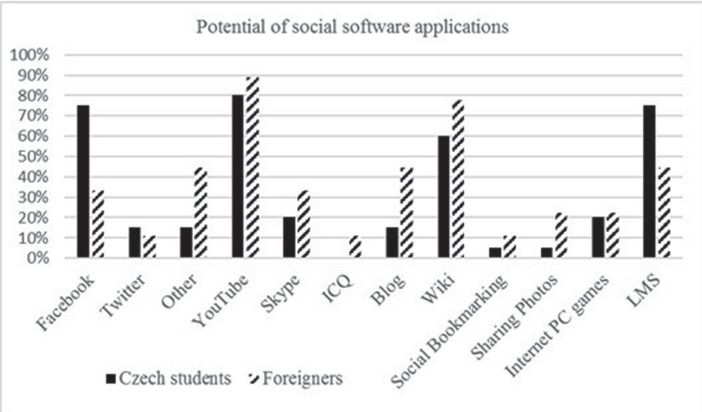


Fig. 6. Potential of social media – comparison Czech a Foreign students view

informal. The strongest media are Facebook and YouTube, which are used daily by 90% of students in case of Facebook and by nearly 80% in case of YouTube, the rest of respondents use them on weekly basis. The question is, if they are used for study purposes. The answer can be found in the responses in the category *Satisfaction* with individual social media for the defined areas of the process of learning (Communication, Study material repository and Revision with Testing). Facebook fits communication perfectly as it is its main mission, anyway. However, students found Facebook dissatisfactory for study material storing and for testing. YouTube has gained incredibly strong position in formal and informal learning. As for LMS, number of students using Blackboard on daily basis has increased which reflects university policy to guarantee support via e-course. Utilization of Wiki has significantly increased, there are 80% users on weekly basis despite the fact that a great deal of teachers are not fans of this platform.

Figure 4 illustrates findings from the latest survey on *Awareness* of social media. The graph is constructed on comparison of Czech and Foreign students' responses. YouTube and Skype reach 100% in both groups. Utilization of Blogs reaches 100% in foreign students and utilization of Social Bookmarking is quite high, both findings support the importance of Blog and Social Bookmarking highlighted in foreign studies. In this case we can see the difference between our scene and scene abroad. Another difference can be seen in ICQ, for foreign students it is 'big unknown', all Czech students know this app. Last comment relates to utilization of LMS; we have to admit a kind of discrepancy in gained responses. Less than half of foreign students marked that they knew LMS, however worked in the e-courses during semester.

As for *Satisfaction,* Twitter, Instagram, WhatsApp (WA), we included in the latest survey see Fig. 5. The 1–5 point scale was applied, Findings bring a colourful picture where level of satisfaction with individual applications differs between Czech and foreign students quite a lot. Czech respondents are quite moderate. As for foreign students, they use the whole scale from 1 to 5. Czech and foreign students are satisfied with YouTube most. Findings relating to Facebook and Twitter social nets are comparable in both groups, surprisingly level of satisfaction is moderate not high, as one might expect and the same refers to Skype. Czech students do not use Blogs as much as foreign students but they are much more satisfied with them then foreign students. Foreign students value pictorial media like Instagram and various kinds of sharing photos.

Students' view of the *Potential* of social media for study purposes shows valuable findings. The highest positions reach YouTube, Wiki in both Czech and foreign students. 90% of foreign students and 80% of Czech students can see potential in the YouTube platform. Three quarters of Czech students can see potential in Facebook in comparison to only 30% of foreign students. Another discrepancy is also perceived in potential of a teaching/learning tool LMS. Three quarters of Czech students can see potential in this social media in comparison to 45% of foreign students. No other application reaches the threshold of 50%.

5 Conclusion

Development of social media and their functionalities never stagnate which imposes challenge to practitioners and researchers. Study climate is changing. To keep the pace with the development of new media and at the same time not to lose unifying view of the social media scene a Social media Acceptance Scheme was designed. This scheme is adaptable; it can be easily modified to fit the purposes of social media research especially in university setting. Students' perspective creates the red thread going through the long-term research: how students perceive, use and value social media. Research targets vary, however awareness of social media platforms represents a kind of continuum being of key importance when considering and creating a digital strategy in the process of education at various levels of its system.

Further development of the research we see in utilization of Unified Theory of Acceptance and Use of Technology (Venkatesh et al. 2003).

Acknowledgement. The paper is supported by the project SPEV 2020 at the Faculty of Informatics and Management of the University of Hradec Kralove, Czech Republic. In addition, the authors thank Anna Borkovcová for her help with the project.

References

Anderson, P.: Web 2.0 and beyond: principles and technologies. Choice Rev. Online **50**(07) (2012). https://doi.org/10.5860/CHOICE.50-3893

Bolton, R.N., et al.: Understanding Generation Y and their use of social media: a review and research agenda. J. Serv. Manage. **24**(3), 245–267 (2013). https://doi.org/10.1108/09564231311326987

Černá, M.: Development of acceptance and utilization of social software applications in higher and further education. In: INTED2014 Proceedings, Presented at the 8th International Technology, Education and Development Conference, pp. 4300–4307 (2014)

Cerna, M.: Potential of social media not only in collaborative learning: expectations and reality case study. In: Cheung, S.K.S., Kwok, L.-f., Yang, H., Fong, J., Kwan, R. (eds.) ICHL 2015. LNCS, vol. 9167, pp. 137–148. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-20621-9_11

Černá, M.: Utilization of social applications as a way of student's engagement into a learning process. Adv. Sci. Lett. **23**(4), 2889–2893 (2017). https://doi.org/10.1166/asl.2017.7595

Černá, M., Borkovcová, A.: Blended learning concept in selected tourism management e-courses with focus on content development including recommender system. In: Cheung, S.K.S., Kwok, L.-f., Kubota, K., Lee, L.-K., Tokito, J. (eds.) ICBL 2018. LNCS, vol. 10949, pp. 175–187. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-94505-7_14

Cerna, M., Borkovcova, A.: Enhancing students' involvement in the process of education through social applications. In: Uskov, V.L., Howlett, R.J., Jain, L.C., Vlacic, L. (eds.) KES SEEL-18 2018. SIST, vol. 99, pp. 185–193. Springer, Cham (2019). https://doi.org/10.1007/978-3-319-92363-5_17

Cerna, M., Lin, C.-J.: Students' language needs analysis as a motivation mover - Czech and Taiwanese case study. In: Popescu, E., Hao, T., Hsu, T.-C., Xie, H., Temperini, M., Chen, W. (eds.) SETE 2019. LNCS, vol. 11984, pp. 99–108. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-38778-5_13

Černá, M., Poulová, P.: Role of social media in academic setting awareness, utilization and willingness. In: 2013 IEEE 11th International Conference on Emerging eLearning Technologies and Applications (ICETA), pp. 59–62. IEEE (2013)

Černá, M., Poulová, P., Draessler, J.: Software applications in academic setting awareness and satisfaction. In: 2011 14th International Conference on Interactive Collaborative Learning, pp. 525–531. IEEE (2011)

Černá, M., Poulová, P., Sláma, M.: Utilization of web portals at Kazakh and Czech University-comparative study. In: Proceedings of the 13th European Conference on e-Learning ECEL 2014, pp. 113–120 (2014)

Černá, M., Poulová, P., Žumárová, M.: Utilization of learning management systems & social networking systems not only in the process of education. In: Recent Researches in Communications, Electrical & Computer Engineering. Presented at the Proceedings of the 10th WSEAS International Conference on Communications, Electrical & Computer Engineering, and 9th WSEAS International Conference on Applied Electromagnetics, Wireless and Optical Communications (ACELAE 2011), pp. 154–159 (2011)

Černá, M., Svobodová, L.: Utilization and expected potential of selected social applications in university setting – a case study. In: Uskov, V.L., Howlett, R.J., Jain, L.C. (eds.) SEEL 2017. SIST, vol. 75, pp. 371–380. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-59451-4_37

Černá, M., Svobodová, L., Haviger, J.: Perception and utilization of Facebook by university students: case study. In: Presented at the European Conference on e-Learning, Academic Conferences International Limited, pp. 105–113 (2017). https://search.proquest.com/openview/0ea99a9d87ccee2f8e104b4ae589b60c/1?pq-origsite=gscholar&cbl=1796419. Accessed 25 July 2020

Cheung, S.K.S.: Distance-learning students' perception on the usefulness of open educational resources. In: Cheung, Simon K.S., Kwok, L.-f., Ma, Will W.K., Lee, L.-K., Yang, H. (eds.) ICBL 2017. LNCS, vol. 10309, pp. 389–399. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-59360-9_34

Cheung, S.K.S.: The open textbooks for Hong Kong: from conceptualization to implementation. In: Cheung, S.K.S., Kwok, L.-f., Shang, J., Wang, A., Kwan, R. (eds.) ICBL 2016. LNCS, vol. 9757, pp. 150–160. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-41165-1_14

Clement, J.: Social media - Statistics & Facts. Statista (2020). https://www.statista.com/topics/1164/social-networks/. Accessed 22 Jan 2020

Czech Statistical Office: Czech Statistical Office Homepage. Czech Statistical Office (2020). https://www.czso.cz/csu/czso/home. Accessed 2 Feb 2020

Eurostat: Individuals using the internet for participating in social networks. Individuals using the internet for participating in social networks (2016). https://ec.europa.eu/eurostat/tgm/table.do?tab=table&init=1&language=en&pcode=tin00127&plugin=1. Accessed 2 Feb 2020

Foreman, C.: 10 Types of Social Media and How Each Can Benefit Your Business. 10 Types of Social Media and How Each Can Benefit Your Business (2017). https://blog.hootsuite.com/types-of-social-media/. Accessed 8 June 2020

Greenhow, C., Lewin, C.: Social media and education: reconceptualizing the boundaries of formal and informal learning. Learn. Media Technol. **41**(1), 6–30 (2016). https://doi.org/10.1080/17439884.2015.1064954

Internet World Stats: Internet World Stats (2020). https://www.internetworldstats.com/surfing.htm. Accessed 2 Feb 2020

Kemp, S.: Digital 2019: Global Internet Use Accelerates, 30 January 2019. https://wearesocial.com/blog/2019/01/digital-2019-global-internet-use-accelerates. Accessed 2 Feb 2020

Lorenc, J.: Social Media stats and demographics in the Czech Republic (2018). https://www.linkedin.com/pulse/social-media-stats-demographics-czech-republic-jakub-lorenc. Accessed 2 Feb 2020

Osch, W.V., Coursaris, C.K.: A meta-analysis of theories and topics in social media research. In: 2015 48th Hawaii International Conference on System Sciences. Presented at the 2015 48th Hawaii International Conference on System Sciences (HICSS), HI, USA, pp. 1668–1675. IEEE (2015). https://doi.org/10.1109/HICSS.2015.201

Pew Research Center: Pew Research Center (2020). https://www.pewresearch.org/topics/social-media. Accessed 2 Feb 2020

Schivinski, B., Christodoulides, G., Dabrowski, D.: Measuring consumers' engagement with brand-related social-media content: development and validation of a scale that identifies levels of social-media engagement with brands. J. Advert. Res. **56**(1), 64–80 (2016). https://doi.org/10.2501/JAR-2016-004

Schroeder, A., Minocha, S., Schneider, C.: The strengths, weaknesses, opportunities and threats of using social software in higher and further education teaching and learning. J. Comput. Assist. Learn. **26**(3), 159–174 (2010)

Schwartz, B., Johnson, B.: The Paradox of Choice: Why More Is Less, Revised Edition. Philosophers Notes, p. 6 (2013)

Selwyn, N.: The use of computer technology in university teaching and learning: a critical perspective. J. Comput. Assist. Learn. **23**(2), 83–94 (2007a)

Selwyn, N.: Web 2.0 applications as alternative environments for informal learning-a critical review. In: Paper for CERI-KERIS International Expert Meeting on ICT and Educational Performance, vol. 16, p. 17 (2007b)

Selwyn, N.: Education and Technology. Key Issues and Debates. Bloomsbury, London (2017)

Statista: Number of social media users worldwide from 2017 to 2025. Statista (2019). https://www.statista.com/statistics/278414/number-of-worldwide-social-network-users/. Accessed 2 Feb 2020

Statista: Social media - Statistics & Facts (2020). https://www.statista.com/topics/1164/social-networks/. Accessed 22 Jan 2020

Steyn, R., Millard, S., Jordaan, J.: The use of a learning management system to facilitate student-driven content design: an experiment. In: Huang, T.-C., Lau, R., Huang, Y.-M., Spaniol, M., Yuen, C.-H. (eds.) SETE 2017. LNCS, vol. 10676, pp. 75–94. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-71084-6_10

Tuten, T.L., Solomon, M.R.: Social Media Marketing. Sage, Thousand Oaks (2017)

Usluel, Y.K., Mazman, S.G.: Adoption of Web 2.0 tools in distance education. Procedia Soc. Behav. Sci. **1**(1), 818–823 (2009)

Venkatesh, V., Morris, M.G., Davis, G.B., Davis, F.D.: User acceptance of information technology: toward a unified view. MIS Q. **27**, 425–478 (2003)

Weller, K.: Accepting the challenges of social media research. Online Inf. Rev. **39**(3), 281–289 (2015). https://doi.org/10.1108/OIR-03-2015-0069

Weller, K.: Trying to understand social media users and usage: the forgotten features of social media platforms. Online Inf. Rev. **40**(2), 256–264 (2016). https://doi.org/10.1108/OIR-09-2015-0299

Weller, K., Dornstädter, R., Freimanis, R., Klein, R.N., Perez, M.: Social software in academia: three studies on users' acceptance of web 2.0 services (2010)

Exploring the Effect of Multichannel Multimodal Learning Environment on Student Motivation and Self-efficacy

Jun Xiao[1(✉)], Tzu-Han Lin[2], and Hong-Zheng Sun-Lin[2]

[1] Shanghai Engineering Research Center of Open Distance Education, Shanghai Open University, No. 288 Guoshun Road, Shanghai 80305, CO, China
ecnuxj2003@163.com

[2] Wisdom Garden Education Research Centre, Kowloon, Hong Kong, China

Abstract. Many multichannel multimodal learning (MML) spaces have been established since the current generation of learners has been getting used to a multichannel multimodal way of experiencing their daily environment. To evaluate the effectiveness of MML space implementation, the present study designed as intervention research examining learning itself the change of motivation and self-efficacy of 48 students at Shanghai Open University before and after their experiencing learning in an MML environment. The difference between students' self-efficacy/intrinsic learning motivation before and after the MML intervention was shown significant (p-value = 0.01535/0.007103), but the difference in students' extrinsic learning motivation was not (p-value = 0.2781). The results pointed out that MML environments enhance students' perceived values of learning itself (compared with its extraneous values); more MML interactions make learning itself more valuable to students, which is evidence supporting the implementation of MML at institutions for learning promotion and teaching quality enhancement.

Keywords: Multichannel multimodal learning · Self-efficacy · Intrinsic motivation · Extrinsic motivation · Learning environment

1 Introduction

Given the rising popularity of the Internet and Information Technology, the current generation of students has got used to adopting a multichannel and multimodal way of experiencing their daily environment. A course that primarily delivers one-mode experience (e.g. text-based course content delivery) or is accessible on one channel (e.g. only in classrooms) is not sufficient. In response to the situation, a concept of multichannel multimodal learning (MML) environments is developed. Modality refers to a way of experiencing course content such as text (i.e. a mode of vision), videos (i.e. a mode of audio and vision), or lectures (i.e. a mode of audio, vision, and likely proprioceptive interaction). Multimodality is a set of different modalities of course content (Moreno and Mayer 2007). Channel refers to a medium for supporting a way of experiencing course

L.-K. Lee et al. (Eds.): ICTE 2020, CCIS 1302, pp. 164–175, 2020.
https://doi.org/10.1007/978-981-33-4594-2_14

content, such as classrooms, computers, or mobile phones. A multichannel learning space permits learners to access the same modality of course content through a variety of channels (Mukhopadhyay and Parhar 2001). An MML environment is an integrated platform where learners can access course content via attending in person, mobile apps, websites, and so on, and course content can be experienced in a set of different modes to suit the varying learning styles of learners.

However, though an MML space theoretically benefits students and is assumed to be able to promote learning, the corresponding evaluation and empirical evidence are missing in the literature. Many studies have examined how other learning spaces (e.g. MOOCs or traditional classrooms) promote learning by measuring learning motivation including self-efficacy (i.e. the expectation component), intrinsic and extrinsic motivation (i.e. the value component). Learners' learning motivation and self-efficacy have thus been regarded as a factor in learning space evaluation (Pintrich and Schunk 1996). Therefore, to fill the mentioned research gap, the present study conducted intervention research on 48 students at Shanghai Open University to explore the changes in their self-efficacy, intrinsic motivation, and extrinsic motivation after they experienced learning in an MML environment. The three variables of interest as indicators for the learning space evaluation will be measured and analyzed in order to answer the research question about whether an MML space enhances student motivation and self-efficacy that promote learning.

2 Literature Review

2.1 Multichannel Multimodal Learning Environment

Learning spaces used to be defined as "brick and mortar" educational institutions where course content was limited to be delivered synchronously (Graetz 2006). However, given the increasing popularity of the Internet, the landscape of course content delivery has been shaped; concepts such as e-learning and distance learning were developed. These concepts represent a form of learning programs that present physical classroom-based instructional content over the Internet. However, there still have been mixed results on which way (i.e., traditional classroom or e-learning) is better for students. Some studies showed e-learning is more likely to attract learners to experience and participate in learning activities than traditional learning (Klesius et al. 1997), some studies also found that learners are substantially less likely to complete online courses compared with face-to-face traditional ones (Carpenter et al. 2004; Xu and Jaggars 2011; Zavarella 2008), and others also showed that correlation between learner academic performance and the two learning spaces is not significant (Thomas 2001).

In fact, the mixed results are not surprising or unexpected since learners represent different generations, different personality types, different learning styles and have different preferences for ways of course content delivery (Graham et al. 2005). According to Gardner (2011), intelligence is not a singular entity but is made up of multiple entities in different proportions used by individuals to understand and to learn about the world, which thus influences how people learn and leads to individual preferences in learning situations. In other words, there is no absolute one type of learning space that is suitable to all types of learners; instead, there is a need to use "multichannel" approaches including face-to-face methods and online technologies delivering "multimodal" course content that meet the needs of a wide spectrum of learners and allow them to engage and experience learning in ways where they have preference, interest or ability and are most comfortable (Picciano 2009).

In this context, concepts such as multichannel and multimodal learning environments are developed. Modality refers to a way of experiencing course content such as text or live streaming; multimodality is a set of different modalities of the same course content (Moreno and Mayer 2007). Channel refers to a device serving as a medium for supporting ways of experiencing course content, such as computers or mobile phones; multichannel learning space permits learners to access the same modality of course content through a variety of channels (Mukhopadhyay and Parhar 2001). Accordingly, the present study defines an MML environment as a multichannel access platform where learners can access to learning via attending in person, mobile apps, websites, and so on, and course content is made and delivered available across a large range of modes to suit the varying learning styles of learners (see Fig. 1). MML will allow learners to find the right mix for themselves out of all the possibilities in learning, no matter if they are offline or online, synchronous, or asynchronous.

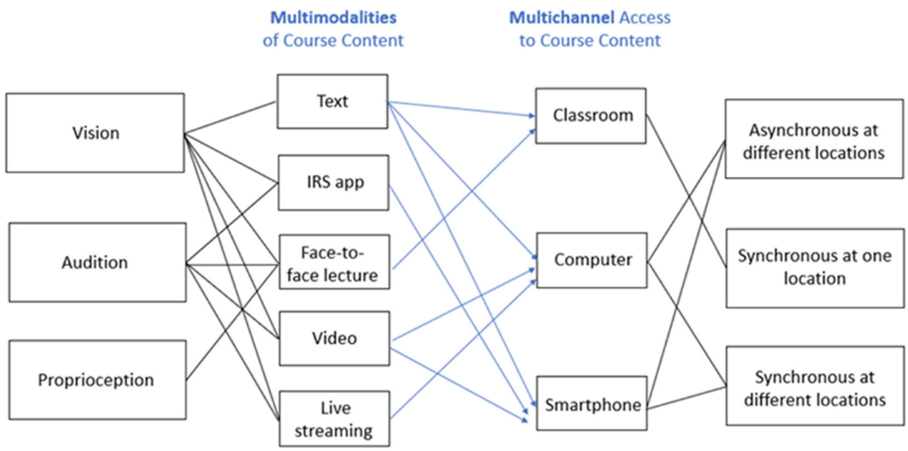


Fig. 1. MML environment integrating multi-modalities and multi-channels

2.2 Learning Motivation and Self-efficacy

Learning motivation has always been regarded as an important factor affecting learning behavior (Pintrich and Schunk 1996) as well as an indicator to measure the quality of teaching (Lee et al. 2010). From the perspective of learning motivation theory (Pintrich and De Groot 1990), learning motivation that can be applied to predict learning behavior and measure teaching quality is mainly composed of two parts: the expectancy component and the value component.

The expectancy component refers to learners' individual belief in and self-judgment on whether they can achieve their learning goals, which is identified as self-efficacy. Self-efficacy was found to be highly correlated with learning motivation (Banfield and Wilkerson 2014). Many empirical studies show that (Zimmerman 1997; Schunk et al. 1987) when students' learning motivation (as an independent variable) increases/decreases, their self-efficacy (as an outcome variable) also increases/decreases; when students' self-efficacy (as an independent variable) increases/decreases, their learning motivation as an outcome variable also increases/decreases. For example, students with high self-efficacy in subject A may not have the same self-efficacy in subject B; because of this specificity, self-efficacy is regarded as an indicator of learning behavior in different contexts (e.g. different subjects or learning settings). Moreover, Multon et al. (1991) conducting a meta-analysis on 39 studies, pointed out that self-efficacy as the expectancy component of learning motivation did have a significant impact on learning behavior to which teaching quality is correlated. Therefore, in terms of learning behavior and teaching quality evaluation, self-efficacy is an important indicator.

The value component can be divided into intrinsic motivation (internal goal orientation) and extrinsic motivation (external goal orientation). Intrinsic motivation means that conducting activities is driven by internal rewards that come from activities (such as learning), such as challenge seeking, so there is no need for extraneous incentives. Activities themselves are the major driving force (Husman and Lens 1999). Intrinsic motivation was found correlated to and considered to have a positive impact on learning behavior (Pintrich and Schunk 1996). On the contrary, extrinsic motivation refers to learning motivation driven by external factors. Learners are not interested in learning itself but are motivated by derivative values (e.g. praise or admiration) from its results (e.g. good performance) (Brophy 2008). Similarly, extrinsic motivation was also found positively associated with learning behavior (Deci et al. 1999). Therefore, in terms of learning behavior measurement and prediction, both intrinsic and extrinsic motivations are important indicators.

2.3 Learning Space Evaluation on Motivation and Self-efficacy

Learning space as a carrier of teaching and learning activities affects learning behavior and teaching quality. Theoretically, as indicators to learning behavior prediction and teaching quality measurement, self-efficacy (the expectation component), intrinsic and extrinsic motivation (the value component) can also be used to evaluate the effectiveness of learning spaces on learning behavior and teaching quality. That is to say, if learning behavior and teaching quality is affected by a learning space, this impact can be reflected by the change of learners' learning motivation and self-efficacy.

Therefore, many studies have explored the correlation between learning motivation and self-efficacy and learning spaces as evaluation research. Lepper (1985) and Zhang et al. (2001) found that virtual learning spaces where learning activities were conducted on computers can make learning more intriguing (compared with that conducted in face-to-face ways) and increase the intrinsic motivation of learning. However, this result may be attributed to the curiosity for technology given that virtual learning was novel at the time. Mei and Hong (2011) established an online asynchronous learning space, where students can interact with each other through discussion forums and message boards.

Designed as a learning space evaluation, the study found that the space did not increase learners' intrinsic motivation but extrinsic motivation. The learners were found motivated to learn by a sense of achievement derived from publicly helping others solve exam and quiz problems, sharing good ideas and thoughts with others, and so on. Following the research of Lepper (1985), Zhang et al. (2001), and Mei and Hong (2011), more and more learning space evaluation studies (Sun and Rueda 2012; Alqurashi 2016; Valencia-Vallejo et al. 2019) have used learning motivation and self-efficacy as variables of interest to measure the effectiveness of learning spaces on learning behavior and teaching quality.

Most studies using learning motivation and self-efficacy for learning space evaluation in literature focused on single-channel (where course content is only accessible through one channel such as classrooms or websites) single-modal (where course materials are only delivered in one mode such as texts or videos) spaces. However, research on evaluating MML environments with learners' motivation and self-efficacy is still limited. Therefore, in response to the research gap, this study will focus on MML spaces and explore their effect on students' intrinsic and extrinsic motivation and self-efficacy which reflects the spaces' impact on learning behavior and teaching quality.

3 Methods

3.1 MML Environment at Shanghai Open University

In order to meet the needs of its diverse student body, Shanghai Open University has built MML environments where course content can be delivered in multimodal forms (e.g. text materials, videos, live streaming) across multichannel accesses to learning (e.g. classrooms on campus, LMS websites, LMS mobile apps, distance, and on-site IRS). For example, a class can be delivered in classrooms (a channel) as a face-to-face lecture (a mode of vision, audition, and proprioception) and online through live streaming (a mode of vision and audition) on LMS websites (another channel) at the same time. Instructors and learners can also interact with each other through IRS mobile apps (another channel with a mode of vision and audition). And the class then can be recorded as a video (a mode of vision and audition) uploaded on LMS websites which learners can access via mobile phones as well. The MML environments support instructors and learners from the phase of course preparation through that of interaction during class to that of supervision after class.

In the present case, a live-broadcasting platform plays an intermediary role that bonds distance learners and on-site instructors and learners together with a chat room for interactive texting and an instant audio/video call system, mimicing real-time communication—which makes options of different modes and channels available for learners. Interactions of different modes (including texts in the chat room) on live streaming would be recorded and uploaded on LMS platforms where learners and instructors can watch the videos and interact asynchronously on the LMS forum. Moreover, not only can the same lectures be delivered in different ways (i.e. multimodal forms), but also learners are allowed to switch between different channels (e.g. they can have a flexible combination of synchronous and asynchronous learning via different devices). Therefore, MML learning experience is realized. The channels and modes the instructor used is shown in Table 1 below, indicating a high level of MML experience in the learning environment.

Table 1. Mode-and-channel-usage pattern of the present case

Channels used		Modes used
Classroom, Computer, & Smartphone	The instructor delivered course content in classroom and on a live-broadcasting platform. The live-broadcasting content would be recorded and uploaded on LMSs. Moreover, students either in classroom or not can interact with the instructor synchronously via an IRS app	Text (vision), Face-to-face lecture (vision, audition, & proprioception), Video (vision & audition), Live streaming (vision & audition), IRS app (vision & audition)

3.2 Research Design

The present study conducted intervention research where an MML space at Shanghai Open University is considered an intervention; in order to explore the effect of the intervention on the variables of interest (i.e., student learning intrinsic motivation, extrinsic motivation, and self-efficacy), the differences between the variables before and after experiencing learning in an MML space were analyzed.

48 students at Shanghai Open University consented to participate in the study. They experienced learning without the intervention in the first half of the semester and with the intervention (i.e. the MML space; see Sect. 3.1) in the second half of the semester. Their self-efficacy, intrinsic motivation, and extrinsic motivation with and without the intervention were surveyed through a questionnaire. The questionnaire was designed upon Lane and Lane (2001)'s self-efficacy survey and Duncan and McKeachie (2005)'s learning motivation survey and consisted of 14 items using a 5 point Likert scale. There were 6 items for self-efficacy, 4 items for intrinsic motivation, and 4 items for extrinsic motivation. Cronbach's alpha for the self-efficacy items is 0.92 (shown in Table 2), that for the intrinsic motivation items is 0.85 (shown in Table 3), and that for the extrinsic motivation items is 0.84 (shown in Table 4. According to Kline (1999), the results indicated that each group of items has good internal consistency (Cronbach's alpha of > 0.80), and the questionnaire is reliable.

Moreover, a statistical power analysis is conducted to make sure the sample size is sufficient. According to Cohen (1988), the power is set to 0.8, the significant level is set to 0.05, and the Cohen's d is set to 0.8 (i.e. large effect size); finally, the minimum sample size is determined to be 20.033, than which the sample size of the present study is higher.

Table 2. Result of reliability analysis on self-efficacy items

raw_ alpha	std.alpha	G6 (smc)	average_r	S/N	ase	mean	sd	median_r	lower	upper
0.92	0.92	0.92	0.65	11	0.018	4.3	0.7	0.64	0.88	0.95

Table 3. Result of reliability analysis on intrinsic motivation items

raw_ alpha	std.alpha	G6 (smc)	average_r	S/N	ase	mean	sd	median_r	lower	upper
0.85	0.86	0.83	0.6	6.1	0.034	4.3	0.7	0.61	0.79	0.92

Table 4. Result of reliability analysis on extrinsic motivation items

raw_ alpha	std.alpha	G6(smc)	average_r	S/N	ase	mean	sd	median_r	lower	upper
0.84	0.83	0.81	0.55	5	0.037	3.5	1.2	0.53	0.76	0.91

These three variables of interest as indicators for learning space evaluation will be measured and analyzed in order to answer the research question below:

Does an MML space enhance student motivation and self-efficacy that encourage learning behavior?

Given the research question, three hypotheses are developed below:

H01: Students' self-efficacy after the intervention of being exposed to an MML environment is "less than or equal to" that before the intervention.

(Ha1: Students' self-efficacy after the intervention of being exposed to an MML environment is "higher than" that before the intervention.)

H02: Students' intrinsic motivation after the intervention of being exposed to an MML environment is "less than or equal to" that before the intervention.

(Ha2: Students' intrinsic motivation after the intervention of being exposed to an MML environment is "higher than" that before the intervention.)

H03: Students' extrinsic motivation after the intervention of being exposed to an MML environment is "less than or equal to" that before the intervention.

(Ha3: Students' extrinsic motivation after the intervention of being exposed to an MML learning environment is "higher than" that before the intervention.)

4 Data Analysis and Results

To test the hypotheses, the present study conducted a one-tailed t-test to analyze mean differences between each of students' self-efficacy, intrinsic motivation, and extrinsic motivation without the intervention and that with the intervention.

In the scope of descriptive statistics, self-efficacy, intrinsic motivation, and extrinsic motivation in the second half of the semester (i.e., after the intervention) are higher than that in the first half of the semester (i.e., before the intervention), see Fig. 2. Without experiencing learning in the MML environment, the mean self-efficacy of students is 4.347 (out of 5), the mean intrinsic motivation is 4.333, and the mean of extrinsic motivation is 3.469. On the other hand, with the intervention, the mean self-efficacy is 4.639, the mean intrinsic motivation is 4.661, and the mean extrinsic motivation is 3.630.

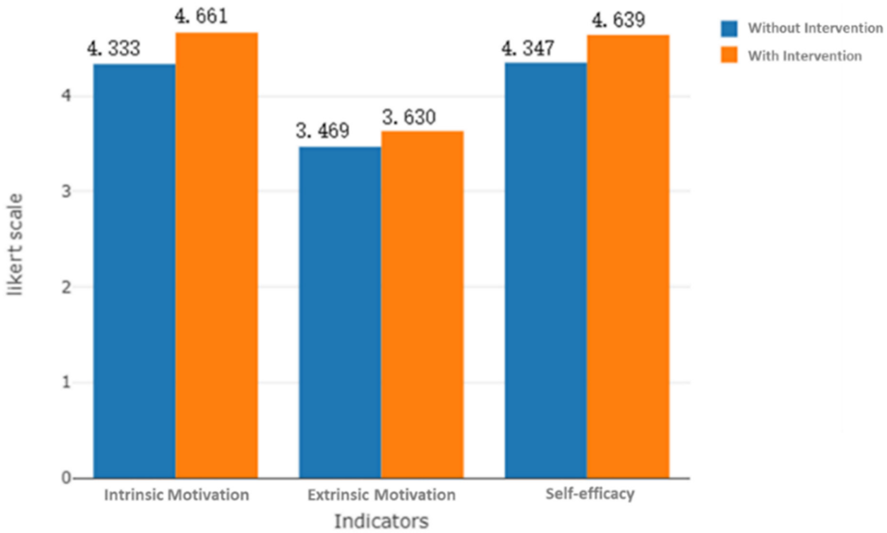


Fig. 2. Comparing students' self-efficacy, intrinsic motivation, and extrinsic motivation before intervention (blue bars) with that after the intervention (orange bars) (Color figure online)

In order to examine whether the differences are statistically significant (more specifically, whether students' self-efficacy, intrinsic motivation, and extrinsic motivation after the intervention of being exposed to the MML environment are "higher than" that before the intervention—i.e., rejecting the null hypotheses), a one-tailed t-test is conducted, and the results are shown in Table 5. In terms of students' self-efficacy, the difference is significant (p-value $= 0.01535 < 0.05$); in terms of students' intrinsic motivation, the difference is significant (p-value $= 0.007103 < 0.05$). However, in terms of students' extrinsic motivation, the difference is not significant (p-value $= 0.2781 > 0.05$).

Therefore, according to the results which reject the first two null hypotheses (i.e., H01 & H02), but fail to reject the third hypothesis, the present study concludes that an MML space generally is capable of enhancing student motivation and self-efficacy that encourage learning behavior; although the improvement on extrinsic motivation is not statistically significant, in the scope of descriptive statistics, its change is still positive (at least not negative) after intervention implementation.

Table 5. One-tailed t-test on the differences in students' self-efficacy, intrinsic motivation, and extrinsic motivation between the second (i.e., with the intervention) and first half of the semester (i.e., without the intervention)

Indicators	Status	N	Mean	t-value	p-value
Self-efficacy	Without the intervention	48	4.347	2.1947	0.01535*
	With the intervention	48	4.639		
Intrinsic motivation	Without the intervention	48	4.333	2.5003	0.007103**
	With the intervention	48	4.661		
Extrinsic motivation	Without the intervention	48	3.469	0.59065	0.2781
	With the Intervention	48	3.630		

$p < 0.05$, **$p < 0.01$.

5 Discussions

According to the results, there are some points worth discussion. First of all, MML environments were found able to improve students' self-efficacy and intrinsic motivation, but not extrinsic motivation. This indicates that creating an MML space can add "values on learning itself" but not create additional "instrumental value of learning." For example, a student who watches course videos because she personally believes they are valuable for her chosen career is extrinsically motivated to learn. Being able to watch the videos through different channels or assimilate the same content in different modes will not affect the extrinsic motivation related to her career plan. On the other hand, since both learning self-efficacy and intrinsic motivation are about learning itself, creating an MML space can enhance the enjoyment of learning itself (i.e., increasing intrinsic motivation) and help overcome obstacles to learning itself (i.e., increasing self-efficacy). Students experiencing learning in the space will feel more efficacious and be motivated intrinsically to learn.

Second, in the scope of descriptive statistics, the participants' self-efficacy and intrinsic motivation before the intervention were high (i.e. 4.347 and 4.333 on a 5-point Likert scale for self-efficacy and intrinsic motivation respectively). This points out a potential lack of external validity of the present study, which may be attributed to participants' self-selection bias that students consenting to make their learning-related behaviors be measured are those who already were confident in or enjoyed learning and thus would not mind their learning process to be analyzed. Moreover, it is also plausible that students consenting to participate in the study are those who were open-minded to studying in different spaces (including different modes and channels of learning), which may make the MML space have fewer negative effects on the sample than general learners.

Finally, since motivation and self-efficacy are not only two factors that promote learning, other factors such as self-regulation, self-determination, learning engagement should be explored in future works in order to more comprehensively evaluate the effect of MML environments on learning as a whole. Moreover, the factors were measured by questionnaires, which only generated self-reported data. Given that MML is not only about inputs (e.g. multichannel multimodal course content) to learning, but also about outputs of learning (e.g. multichannel multimodal data collection of students' learning), it is worth considering measuring the factors by collecting multichannel multimodal data in an MML space. For example, self-regulation as a latent factor related to learning behavior was found measured through multichannel multimodal data in literature. Malmberg et al. (2019) conducted a study collecting multichannel multimodal data (including physiological data, video observations, and facial recognition data) to measure the types of regulation of learning. Wise and Hsiao (2019) examined students' regulation of listening and speaking by multichannel multimodal data such as click-stream data and manually coded post content for argumentation (Azevedo and Gašević 2019). These studies pointed out that measuring learning behavior through multichannel multimodal output data of learning is promising and feasible for future works.

6 Conclusions

One of the core design principles of learning environments is to create settings strengthening the interaction between learners, course content, and instructors for enhancing learning motivation and self-efficacy. Thus, learners' motivation and self-efficacy are usually measured to evaluate the effectiveness of learning space implementation. The present study evaluated an MML environment by conducting an intervention study on the change of learning motivation and self-efficacy of students at Shanghai Open University. The results showed that learners' self-efficacy and intrinsic motivation can be enhanced after experiencing MML interactions with course content. This indicated that building learning spaces helping lecturers deliver MML experience is not only to meet learners' needs (as the current generation of students has been getting used to a multichannel multimodal way of interacting with daily information), but also to make learning a more inherently interesting or enjoyable work that students feel more confident to do. The present study on the effectiveness of MML environments provided evidence supporting the popularization of MML spaces at educational institutions.

Acknowledgment. This paper is supported by China's National General Project granted by China National Office for Education Sciences Planning (Grant No. BCA160053), The Construction and Application of Online Learners' Persona based on Big Data Analysis.

References

Alqurashi, E.: Self-efficacy in online learning environments: a literature review. Contemp. Issues Educ. Res. **9**(1), 45–51 (2016)

Azevedo, R., Gašević, D.: Analyzing multimodal multichannel data about self-regulated learning with advanced learning technologies: Issues and challenges. Comput. Hum. Behav. **96**, 207–210 (2019)

Banfield, J., Wilkerson, B.: Increasing student intrinsic motivation and self-efficacy through gamification pedagogy. Contemp. Issues Educ. Res. **7**(4), 291–298 (2014)

Brophy, J.: Developing students' appreciation for what is taught. Educ. Psychol. **43**(3), 132–141 (2008)

Carpenter, T.G., Brown, W.L., Hickman, R.C.: Influences of online delivery on developmental writing outcomes. J. Dev. Educ. **28**(1), 14–16 (2004)

Cohen, J.: Statistical Power Analysis for the Behavioral Sciences, 2nd edn. Erlbaum, Hillsdale (1988)

Deci, E.L., Koestner, R., Ryan, R.M.: A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation. Psychol. Bull. **125**, 627–668 (1999)

Duncan, T.G., McKeachie, W.J.: The making of the motivated strategies for learning questionnaire. Educ. Psychol. **40**(2), 117–128 (2005)

Gardner, H.: Frames of Mind: The Theory of Multiple Intelligences. Hachette UK, London (2011)

Graetz, K.A.: The psychology of learning environments. Educ. Rev. **41**(6), 60–75 (2006)

Graham, C.R., Allen, S., Ure, D.: Benefits and challenges of blended learning environments. In: Khosrow-Pour, M. (ed.) Encyclopedia of Information Science and Technology, pp. 253–259. Idea Group Inc., Hershey (2005)

Husman, J., Lens, W.: The role of the future in student motivation. Educ. Psychol. **34**(2), 113–125 (1999)

Klesius, J.P., Homan, S., Thompson, T.: Distance education compared to traditional instruction: the students' view. Int. J. Instr. Media **24**(3), 207–227 (1997)

Kline, T.J.: The team player inventory: reliability and validity of a measure of predisposition toward organizational team-working environments. J. Spec. Group Work **24**(1), 102–112 (1999)

Lane, J., Lane, A.: Self-efficacy and academic performance. Soc. Behav. Pers. Int. J. **29**(7), 687–693 (2001)

Lee, J.C.K., Zhang, Z., Yin, H.: Using multidimensional Rasch analysis to validate the Chinese version of the Motivated Strategies for Learning Questionnaire (MSLQ-CV). Eur. J. Psychol. Educ. **25**(1), 141–155 (2010)

Lepper, M.R.: Microcomputers in education: motivational and social issues. Am. Psychol. **40**, 1–18 (1985)

Malmberg, J., Järvelä, S., Holappa, J., Haataja, E., Huang, X., Siipo, A.: Going beyond what is visible: what multichannel data can reveal about interaction in the context of collaborative learning? Comput. Hum. Behav. **96**, 235–245 (2019)

Mei, Z., Hong, J.: Implementing theories in educational psychology in teaching and learning: how teachers facilitate students' learning motivation through the use of e-learning. Hong Kong Teachers' Centre J. **10**, 27–33 (2011)

Moreno, R., Mayer, R.: Interactive multimodal learning environments. Educ. Psychol. Rev. **19**(3), 309–326 (2007)

Mukhopadhyay, M., Parhar, M.: Instructional design in multi-channel learning system. Br. J. Edu. Technol. **32**(5), 543–556 (2001)

Multon, K.D., Brown, S.D., Lent, R.W.: Relation of self-efficacy beliefs to academic outcomes: a meta-analytic investigation. J. Couns. Psychol. **38**(1), 30 (1991)

Picciano, A.G.: Blending with purpose: the multimodal model. J. Asynchronous Learn. Netw. **13**(1), 7–18 (2009)

Pintrich, P.R., De Groot, E.V.: Motivational and self-regulated learning components of classroom academic performance. J. Educ. Psychol. **82**(1), 33 (1990)

Pintrich, P.R., Schunk, D.H.: Motivation in Education: Theory, Research, and Applications. Englewood Cliffs, Prentice Hall (1996)

Thomas, R.: The no significant difference phenomenon: a comparative research annotated bibliography on technology for distance education. IDECC, Montgomery, AL (2001)

Schunk, D.H., Hanson, A.R., Cox, P.D.: Peer-model attributes and children's achievement behaviors. J. Educ. Psychol. **79**(1), 54 (1987)

Sun, J.C.Y., Rueda, R.: Situational interest, computer self-efficacy and self-regulation: their impact on student engagement in distance education. Br. J. Edu. Technol. **43**(2), 191–204 (2012)

Valencia-Vallejo, N., López-Vargas, O., Sanabria-Rodríguez, L.: Effect of a metacognitive scaffolding on self-efficacy, metacognition, and achievement in e-learning environments. Knowl. Manage. E-Learning Int. J. **11**(1), 1–19 (2019)

Wise, A.F., Hsiao, Y.T.: Self-regulation in online discussions: aligning data streams to investigate relationships between speaking, listening, and task conditions. Comput. Hum. Behav. **96**, 273–284 (2019)

Xu, D., Jaggars, S.S.: Online and hybrid course enrollment and performance in Washington State community and technical colleges. Community College Research Center, Columbia University, New York, NY (2011). http://ccrc.tc.columbia.edu/media/k2/attachments/online-hybridperformance-washington.pdf

Zavarella, C.A.: Computer-based instruction and remedial mathematics: a study of student retention at a Florida community college. Doctoral dissertation, Available from ProQuest Dissertations and Theses database (UMI No. ATT 33260369) (2008)

Zhang, J., Li, F., Duan, C., Wu, G.: Research on self-efficacy of distance learning and its influence to learners' attainments. In: Proceedings of the International Conference on Computers in Education, pp. 1510–1517 (2001)

Zimmerman, E.: Building leadership roles for teachers in art education. J. Art Des. Educ. **16**(3), 282–284 (1997)

Establishing a Tripartite Intelligent Reading Platform Connecting Schools, Families, and Students: An Approach with Chinese Characteristics to Promote Students' Reading

Jielin Zeng[1], Yirao Liu[2], Xiaoxiao Huang[3], Ruixiang Gao[4], Hu Mei[5](✉), and Zehui Zhan[6](✉)

[1] School of Geography, South China Normal University, Guangzhou 510631, China
[2] School of Physics and Telecommunication Engineering, South China Normal University, Guangzhou 510006, China
[3] School of Mathematical Sciences, South China Normal University, Guangzhou 510631, China
[4] School of Psychology, South China Normal University, Guangzhou 510631, China
[5] Key Lab for Behavioral Economic Science and Technology, School of Economics and Management, South China Normal University, Guangzhou 510006, China
meihu@m.scnu.edu.cn
[6] School of Information Technology in Education, South China Normal University, Guangzhou 510631, China
zhanzehui@m.scnu.edu.cn

Abstract. Reading is important for students, but despite many efforts made to promote it, Chinese students generally do not spend much time reading. We believe the bottleneck lies in the difficulty of retaining students' reading data, which makes it difficult for teachers and parents to know, manage, and guide students' reading behaviors. Therefore, we establish a tripartite intelligent reading platform connecting schools, families, and students: schools assign reading tasks and give instructions on the platform; students "punch in" on the platform before and after the reading tasks; parents can get to check on their kids' performance through the platform and also keep encouraging and supervising them to read. The platform has now been used in 110 schools in Guangzhou, and its significant effects are proved by large-scale survey statistics. Compared with traditional ways that aim at arousing students' reading interest in the first place but are often eventually defeated by students' massive compulsory homework under China's educational system, our approach using mandatory methods is of great Chinese characteristics. In the beginning, the students might be "forced" to read by family-school collaboration. But gradually, external motivation would be transformed into internal motivation; students are becoming actively engaged in reading. In the future, based on the platform, Chinese governments and research institutes are preparing to establish large databases concerning reading and student's development, which will hopefully make great contributions to educational decision-making and scientific research.

Keywords: Students' reading · Intelligent reading platform · Family-school collaboration · Chinese characteristic · Large database

L.-K. Lee et al. (Eds.): ICTE 2020, CCIS 1302, pp. 176–188, 2020.
https://doi.org/10.1007/978-981-33-4594-2_15

1 Introduction

Reading is an important way for people, especially children and teenagers, to acquire knowledge and increase wisdom. It is also an important way for a country and a nation to develop their spirit and pass on their civilization. This is why many countries around the world attach great importance to the promotion of students' reading and take active measures. Japan, for example, has legislated to ensure students' reading since 1940s, followed by USA in the 1980s and Korea and Russia around the early 2000s [1].

China advocated reading nationwide in the government work report of the state council for the very first time in 2014 and has continued for the next six years [2]. In 2016, the first national file planning for civil reading during the 13th five-year period was released, emphasizing on the highest priority of children [3]. MOE of China called for extracurricular reading half an hour each day for primary school students and an hour each day for middle school students [4]. In the wake of a series of policies, practices to promote students' reading are carried out in full swing across the country, among which are mainly two folds: one is to ensure hardware facilities including enough libraries, books, teachers, and so on; the other is to organize varieties of reading activities to stimulate students' interest. The practices to promote students' reading around the world are nothing more than these two kinds, but measures proved effective in other countries mostly failed under Chinese circumstance. This is largely because in Chinese exam-oriented education system, students are too busy coping with massive compulsory homework and barely have extra time for reading, even though some of them are very fond of it. Likewise, teachers and parents often give way to textbook studying and examination practicing, even though most of them are aware of the fact that extracurricular reading will contribute to improving test scores. Consequently, reading has not yet been able to place on students' agenda.

As to how to promote Chinese students' reading, it is a hard nut to crack. We assume that the reason why traditional approaches which aim at arousing students' reading interest in the first place did not work is that the basic time for students' extracurricular reading was not guaranteed. Organizing reading activities can only stir up students' interest in short-term training, but without consecutive follow-up measures to keep students' eyes on books, aroused reading interest will not last long. Thus, this kind of approaches is unsustainable. To tackle this issue, we must schedule extracurricular reading in students' daily routine with mandatory supervision first before they develop their reading interest. And only when students are constantly engaged in reading for a period of time, no matter forcedly or voluntarily, can they dig out the real fun of reading. Thus, the problem left to be solved is to find a way for teachers and parents can know, manage, and guide students' extracurricular reading. In the following sections, we are going to introduce a tripartite intelligent reading platform that we establish to settle this problem, and its significant effect for promoting students' reading.

2 The Intelligent Reading Platform

According to the above analysis, in our perspective, the key bottleneck we face in the promotion of students' reading lies in the difficulty to obtain and supervise students' reading behaviors. If teachers and parents cannot know about students' extracurricular reading, they will not be able to manage and guide them [5]. Hence, the major goal of this platform is to preserve students' reading behaviors and to enable teachers and parents to monitor students cooperatively and make reading a daily must-do task for students. The platform consists of three terminals: the school terminal, the student terminal, and the family terminal (Fig. 1).

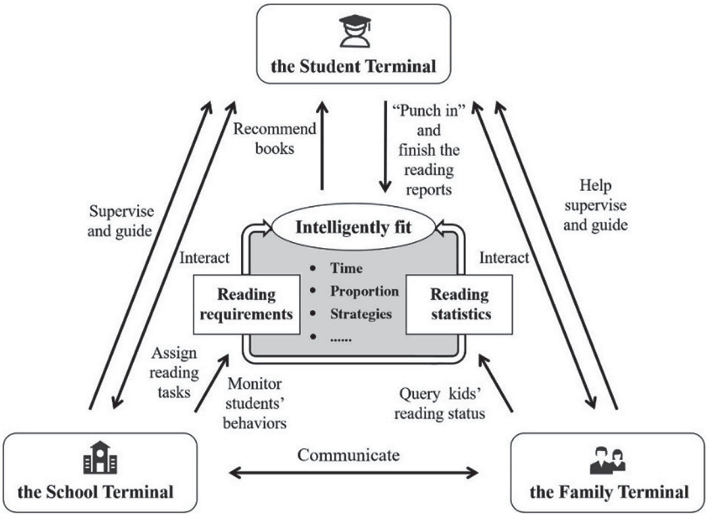


Fig. 1. The structure of the tripartite intelligent reading platform.

2.1 The School Terminal

The main role of the school terminal is to assign reading tasks to students and monitor their behaviors. Through the school terminal, school teachers can deliver various reading instructions, requirements, or guidance to classes or individuals under their charge, such as the total amount of reading required to complete this semester, required book lists, reading ratio requirements for each kind of books, suggested reading strategies and methods, etc. Teachers, school administrators, and leaders in educational departments can query the reading statistics of individuals, groups, classes, and schools within the scope of their different access authorities, including the amount of reading, completion for required books, the proportion of each type of books read, the use of strategies and methods when reading, etc. Also, they can communicate with other people through the platform. It should be noted that not only Chinese teachers can join the platform, but also teachers of other subjects can recommend reading books and launch reading activities related to the subject on the platform.

2.2 The Student Terminal

The main role of the student terminal is for students to select books and complete their reading report. On the student terminal, each student has their own account. After logging on, students need to fill in their personal information including age, grade, gender, reading interest, expectation for future development, etc. According to the students' personal information and current reading statistics and teachers' reading instruction, the system will intelligently recommend books for the student. Notably, in consideration of protecting students' vision, we do not encourage e-reading, so our platform does not provide online reading resources, but informs students of the library or bookstore where they can borrow or purchase recommended printed books instead.

After students get the books, they begin to read them as required, and then log on to the platform regularly to "punch in" and finish the reading reports. The reports mainly include reading time, reading strategies and methods, reading impressions and gains, and/or completing corresponding post-reading test questions assigned by the teachers. In order to encourage students to use the intelligent reading platform, we also set up some incentive mechanism. For instance, students who have completed a certain amount of reading can earn and accumulate points which can be exchanged for small gifts. In this way, students' reading behaviors can be recorded.

At the same time, every student can see his/her reading status in the class on the platform, and students who fall behind will receive a small reminder from the platform encouraging them to catch up.

2.3 The Family Terminal

The role of the family terminal is to facilitate parents to see their kids' reading status, and then help teachers supervise and guide students' extracurricular reading. Through the family terminal, parents can query students' reading data, communicate with the teachers, and interact with their children, such as giving comments and likes to students' works shared after reading.

The establishment of this platform is to solve the problem of data retention of students' reading behaviors, making students' extracurricular reading "knowable, manageable, and guidable" through the joint efforts of both teachers and parents.

3 Evaluation of the Platform's Effect on Promoting Reading

Currently, with the support of policy from the education bureau, the reading platform has been established in more than 100 primary schools in Guangzhou to carry out an experiment of promoting students' reading since the end of 2018. In order to evaluate the effect of the platform, we conducted a survey between experimental schools (E schools) and non-experimental schools (NE schools) in October, 2019.

3.1 Method

The survey mainly includes two aspects: one is the comparison on extracurricular reading between the two types of schools' students, including the awareness of the importance of reading, the investment of time in reading, the reading ratio of different kinds of books, the use of reading strategies, the benefits from reading, etc.; the other is the comparison on the supporting facilities and measures provided and adopted by schools and families between the two groups, such as library resources, reading requirements, reading guidance, etc.

According to the above two aspects of the survey, three questionnaires were designed respectively for students, teachers, and parents. The questionnaires were distributed through the Internet. The students, teachers, and parents of experimental schools and non-experimental schools were randomly sampled. Among them are 12,902 students, 11,222 parents, and 470 teachers from experimental schools; 11,754 students, 10,118 parents, and 567 teachers are from non-experimental schools.

In order to ensure the authenticity of the data collected, the data of some items of this survey had information sources from three or two aspects: students, teachers, and parents, which were summed up in a weighted method in the calculation. For example, the data of the students' question of "how much time do you spend in reading extracurricular books on average every day" will be given a weight of 0.5, and the same question will also be asked for teachers ("how much time do your students spend in reading extracurricular books on average every day") and parents ("how much time does your child spend in reading extracurricular books on average every day"), whose answers are respectively given a weight of 0.25. If there are only two aspects of data (because some questions are not known by all three kinds of respondents), the weight of the most important information source (directly answering the question) is 0.75 and the weight of the auxiliary information source (indirectly reflecting the question) is 0.25. The data listed in the tables in this article are the final outcomes calculated according to the above method. And all the data are from *White Paper on Reading for Primary School Students in Guangzhou* [6].

3.2 Result

Students' Reading. *Average Reading Time.* As shown in Table 1, the number of students who spent more than 30 min in reading every day (reaching the expected standard of MOE) in the experimental schools was 45.94%, while that in non-experimental schools was only 34.45%. We used a 2 × 2 chi-square test to prove that the difference was statistically significant ($\chi^2 = 368.79$, $p < 0.0001$). It shows that the establishment of the platform has substantial effect in promoting students' extracurricular reading.

Table 1. The survey of students' reading time.

Item	How much time do you spend in reading extracurricular books on average every day?			
Options	<15 min	15–30 min	30–60 min	>60 min
E schools	7.10%	46.96%	38.18%	7.77%
NE schools	12.70%	52.86%	28.89%	5.56%

Reading Interests. As shown in Table 2, the majority of primary school students like extracurricular reading. Specifically, the number of experimental schools' students who "strongly like" extracurricular reading is significantly larger than that of non-experimental schools ($\chi^2 = 280.38$, $p < 0.0001$), which means the platform has indeed enhanced students' interest in reading.

Table 2. The survey of students' reading interest.

Item	Do you like reading extracurricular books?			
Options	Dislike	General	Like	Strongly like
E schools	0.98%	18.43%	46.92%	33.68%
NE schools	1.51%	25.17%	49.34%	23.98%

Awareness of Importance of Reading. As shown in Table 3, most of the students believe extracurricular reading is necessary. And the difference between the experimental schools' and the non-experimental schools' students who believe extracurricular reading is "very necessary" is significant, as revealed by a 2 × 2 chi-square test ($\chi^2 = 91.74$, $p < 0.001$).

Table 3. The survey of students' awareness of importance of reading.

Item	Do you think it necessary to read extracurricular books?			
Options	Entirely unnecessary	It doesn't matter	Necessary	Very necessary
E schools	1.78%	6.17%	52.37%	39.68%
NE schools	2.50%	8.10%	55.82%	33.58%

Reading Ratio of Each Kind of Books. As shown in Table 4, although more than three quarters of primary school students prefer reading fairy tale and cartoon books, the number of students in experimental schools who prefer reading traditional cultural classics and Chinese revolutionary stories has increased about 5%, which are favored by the Chinese government.

Reading Strategies and Methods. The use of reading strategies and methods can be seen in Table 5. 2 × 2 chi-square tests demonstrated that students in experimental schools can significantly better understand the passage through skimming ($\chi^2 = 74.55$, $p < 0.001$), read with questions in mind ($\chi^2 = 36.36$, $p < 0.001$), infer the meaning of new words from the context ($\chi^2 = 99.71$, $p < 0.001$), and retell what they read to others after reading ($\chi^2 = 27.54$, $p < 0.01$).

Table 4. The survey of students' reading ratio of each kind of books.

Item	What kind of books do you read the most outside class? (Choose no more than 3 options)							
Options	Novel, Prose, & Poetry	Fairy tale, & Cartoon book	Science fiction, & Suspense	History, & Biography	Science, & technology	Traditional cultural classic	Chinese revolutionary story	News
E schools	36.26%	75.07%	34.39%	24.42%	35.71%	33.75%	11.07%	2.35%
NE schools	35.71%	79.10%	30.02%	22.48%	32.81%	28.20%	7.51%	2.78%

Table 5. The survey of students' reading strategies and methods.

Item	Can you get a general idea of what you read by skimming?		
Options	Yes	Sometimes	No
E schools	45.71%	50.50%	3.80%
NE schools	41.57%	53.22%	5.21%
Item	Can you read with questions in mind?		
Options	Yes	Sometimes	No
E schools	36.48%	58.58%	4.94%
NE schools	34.33%	59.65%	6.02%
Item	Can you infer the meaning of new words from the context while reading?		
Options	Yes	Sometimes	No
E schools	35.30%	57.95%	6.75%
NE schools	30.26%	60.50%	9.23%
Item	Can you retell what you read to others after reading?		
Options	Yes	Sometimes	No
E schools	47.23%	49.14%	3.64%
NE schools	43.75%	51.90%	4.35%

Benefits from Reading. The educational effect of extracurricular reading can be seen in Table 6. 2 × 2 chi-square tests demonstrated that students in experimental schools believe reading significantly improves their Chinese proficiency ($\chi^2 = 11.78, p < 0.01$), betters their understanding of the outside world ($\chi^2 = 60.57, p < 0.001$), helps them to have a more positive mindset ($\chi^2 = 58.52, p < 0.001$), and promotes the quality of their interpersonal relationships ($\chi^2 = 38.31, p < 0.001$).

In conclusion, students in the experimental schools with the intelligent reading platform established became more interested in reading, spent more time in reading, became

Table 6. The survey of students' benefits from reading.

Item	Does extracurricular reading improve your Chinese proficiency?			
Options	No help	A little help	Greater help	Much greater help
E schools	0.95%	23.93%	42.40%	32.72%
NE schools	0.98%	25.86%	42.48%	30.68%
Item	Does extracurricular reading help you understand the outside world better?			
Options	No help	A little help	Greater help	Much greater help
E schools	0.64%	16.39%	42.66%	40.31%
NE schools	0.62%	19.67%	44.21%	35.49%
Item	Does extracurricular reading make you more positive and optimistic?			
Options	No help	A little help	Greater help	Much greater help
E schools	1.67%	18.53%	40.77%	39.03%
NE schools	1.52%	21.13%	43.02%	34.33%
Item	Does extracurricular reading help you get along with others better?			
Options	No help	A little help	Greater help	Much greater help
E schools	2.05%	20.34%	40.03%	37.58%
NE schools	1.74%	22.77%	41.70%	33.79%

more aware of the importance of reading, read more books with better quality and wider range, learned to employ more strategies while reading, and most importantly, gained more benefits through reading.

Reading Support from Schools. *Emphasis from School Leadership.* As shown in Table 7, most school leaders attach great importance to students' extracurricular reading. Specifically, 68.09% of the leaders in the experimental schools has put a lot of emphasis on reading, while only 51.32% leaders in non-experimental schools have done so. The difference was significant, as revealed by a 2 × 2 chi-square test ($\chi^2 = 29.83$, $p < 0.01$), which suggests that the establishment of the platform is conducive to deepening the understanding of school leaders on students' reading, so as to better promote it at the school level.

Table 7. The survey of attitude of school leadership towards students' reading.

Item	What is the attitude of school leadership towards reading?			
Options	Pay no attention	General	Emphasis	High emphasis
E schools	0.21%	6.17%	25.53%	68.09%
NE schools	0.71%	11.29%	36.68%	51.32%

Reading Supervision from School Teachers. As shown in Table 8, the majority of school teachers make requirements and conduct inspection on the books that students read. Concrete measures were taken by 70.29% of the teachers in the experimental schools but only 51.32% teachers in non-experimental schools; the difference was also significant ($\chi^2 = 8.41$, $p < 0.01$). It suggests that the establishment of the platform facilitates the performance of teachers' duties.

Table 8. The survey of teachers' supervision on students' reading books.

Item	Are there any specific requirements from teachers to make sure students choose good books to read, especially the ones recommended by MOE, and are there any measures by teachers to inspect students' reading?			
Options	No requirement, No inspection	Occasional requirement, No inspection	Regular requirement, No inspection	Regular requirement and inspection
E schools	0.53%	8.83%	20.36%	70.29%
NE schools	1.59%	10.72%	25.55%	62.14%

Reading Guidance from School Teachers. As shown in Table 9, the difference between the numbers of teachers who have given concrete instructions to students' reading in the two types of schools was significant as well ($\chi^2 = 119.49$, $p < 0.001$).

Table 9. The survey of teachers' guidance on students' reading strategies and methods.

Item	Do the teachers emphasize the importance of reading strategies and methods and give specific instructions on students' extracurricular reading?			
Options	No emphasis, No instructions	A little emphasis, No instructions	A lot of emphasis, Few instructions	A lot of emphasis, A lot of instructions
E schools	2.54%	13.30%	35.57%	48.59%
NE schools	3.26%	17.94%	35.52%	43.28%

Reading Support from Families. *Parents' Knowledge of Students' Reading.* As shown in Table 10, more than two thirds of the parents know about their children's reading, but the number in experimental schools was still significantly higher, as revealed by a 2 × 2 chi-square test ($\chi^2 = 93.45$, $p < 0.0001$), implying that the intelligent reading platform makes the information of students' extracurricular reading more transparent.

Parents' Attitude Towards Students' Reading. As shown in Table 11, most of the parents are supportive of their children's reading, and still, the number in experimental schools is significantly higher ($\chi^2 = 47.58$, $p < 0.001$), demonstrating the fact that the platform contributes to changing parents' attitudes towards students' extracurricular reading.

Table 10. The survey of parents' knowledge of students' reading.

Item	Do you know about your child's extracurricular reading?			
Options	Know nothing	Know a little	Know a lot	Know very well
E schools	1.93%	23.03%	43.49%	31.55%
NE schools	3.23%	29.56%	41.65%	25.56%

Table 11. The survey of parents' attitude towards students' reading.

Item	What is your attitude towards your child's extracurricular reading?			
Options	No support	No concern	Encourage	Highly encourage
E schools	0.44%	2.60%	46.21%	50.75%
NE schools	0.44%	3.63%	50.08%	45.85%

Number of Books at Home. As shown in Table 12, the number of families in the experimental schools who keep more than 50 books at home reached 61.49%, while that in the non-experimental schools was only 47.55%. A 2 × 2 chi-square test revealed that the difference was significant ($\chi^2 = 397.36$, $p < 0.0001$), which means the platform has helped perfected students' reading environments.

Table 12. The survey of the number of books at home.

Item	How many books do you have in your home (excluding students' textbooks and exercises)?			
Options	<20	<20	<20	<20
E schools	9.36%	29.16%	30.11%	31.38%
NE schools	18.63%	33.83%	26.15%	21.40%

Parents' Ability of Reading Guidance. As shown in Table 13, limited as parents' guidance was, the ratio of parents in the experimental schools who can provide reading guidance for their children was significantly higher when compared with that in the non-experimental schools, as confirmed by a 2 × 2 chi-square test ($\chi^2 = 25.28$, $p <$ 0.01).

Based on the above figures from both teachers and parents, conclusion can be drawn that the establishment of the intelligent reading platform has indeed made students' extracurricular reading more "knowable, manageable, and guidable" and effectively promoted students' reading through home-school collaboration.

Our intelligent reading platform was highly recognized by MOE of China in 2017 [7], and was popularized in *Report on the Work of Guangzhou Government 2019* [8].

Table 13. The survey of parents' ability of reading guidance.

Item	To what extent can you guide your child in extracurricular reading?			
Options	Under educated, No guidance	Limited reading experience, Limited guidance	Limited time and energy, Limited guidance	Able to guide to a great extent
E schools	4.07%	16.39%	40.23%	39.31%
NE schools	5.56%	18.09%	40.38%	35.97%

4 Discussion and Conclusion

4.1 The Role of the Platform

The difficulty in retention of students' extracurricular reading behaviors, which results in the difficulty in educational management of reading, has long existed and remained unsolved. Though there are a great deal of intelligent reading platforms existing, they without exception rely on electronic reading, which is not encouraged for a series of disadvantages especially for young children. Comparatively speaking, our intelligent reading platform is smartly designed, using incentive measures to have students willingly "punch in" every time they read a book. In this way, students' extracurricular reading behaviors can be recorded as data.

The authenticity of students' reading report on our platform is most frequently questioned, for students might exaggerate their reading in order to be praised. But in our sense, this would not be a big problem, because too much exaggeration will be easily spotted by the teachers and this kind of false would not last long. Most importantly, we believe students can naturally develop their interest when engaged in reading for a period of time, as will be discussed later; once they have inner incentive for reading, they will not cheat any more. And to demonstrate our point, the survey we conducted did not use the existing data on the platform but collected data by other means with more than one source of information instead, and the results adequately testify the usefulness of the platform in promoting students' reading.

Moreover, the platform serves as a bridge for parents and teachers to communicate with each other and to make mutual effort to supervise and guide students' reading, which is the most essential factor accounting for the growth of students' reading. As can be clearly seen in the data, the platform has successfully prompted school leaders and teachers to implement reading supervision and surprisingly enhanced parents' ability in reading guidance.

4.2 Wisdom with Great Chinese Characteristics

Admittedly, this kind of approach establishing reading platforms in all schools of a city by administrative order is apparently with great Chinese characteristics, which, in other words, might not be able to borrow to other countries with a different national system. But at the same time, this approach is well adapted to Chinese context. As mentioned in the introduction, nearly everything concerned students in China has to give way to the college entrance examination. Students' extracurricular reading, which seems indirectly related to the examination and lacks certain urgency and coerciveness, is easily marginalized. This is why many approaches proved effective in other countries more often than not failed in China. To tackle this problem, we have to come up with a way also with Chinese characteristics, that is to "force" students to read at the beginning whether students like it or not.

Unique, weird, and even "politically incorrect" as it might be accused, our approach has been proved successful by data. At first, students might be passively forced to read by family pressure and peer comparison, because they were afraid of criticism from teachers and parents if they fell behind in reading statistics on the platform. But gradually, they would be captured by the natural charm of reading and began to fall in love with it. As can be seen in Table 1 and 2, students reading interest has increased. Put it in the perspective of educational psychology, external motivation has transformed into internal motivation. Henceforth, students would seek to read initiatively and independently. And students' reading abilities would be improved as well.

The wisdom behind this approach is the idea with great Chinese characteristics of "Love what you do (read)" instead of "Do (read) what you love", which is widely different from western "interest-oriented" learning belief [9]. This idea is also embodied in the requirement of balanced reading; students are supposed to read as widely as possible, including some salutary books they might not like.

4.3 Lasting Positive Impact

Furthermore, the establishment of our intelligent reading platform has long-lasing positive impact. As reading research is becoming a heated topic in education, psychology, brain science, artificial intelligence, and other fields, the questions as to how to effectively develop students' reading literacy and how to explore and uncover the neural mechanisms of reading and the influence of reading on the all-round development of students have now been major scientific topics drawing international academic attention. Our intelligent reading platform solves the problem of retention of students' reading data, which not only contributes to breaking through the bottleneck in reading educational management, but also is conducive to boosting the scientific research of reading to achieve fruitful results. In the future, based on our platform, governments and research institutes are preparing to build the world's first large database concerning students' reading and brain development, which hopefully make great contributions in providing service for educational decision-making of the governments and reading-related scientific research [10].

Acknowledgement. This study was supported by the Major basic research and applied research projects of Guangdong Education Department (#2017WZDXM004), MOE (Ministry of Education in China) Project of Humanities and Social Sciences (#19YJC880125); National Social Science Foundation of China (#18BGL053).

References

1. Zhang, Q.L.: Comparison and enlightenment of campus reading legislation in foreign studies. Libr. Work Study **08**, 110–113 (2016)
2. State Council of the PRC. http://www.gov.cn/guowuyuan/2020zfgzbg.htm. Accessed 30 Sept 2020
3. Xinhuane. http://www.xinhuanet.com/politics/2016-12/27/c_129421928.htm. Accessed 30 Sept 2020
4. MOE of PRC. http://www.moe.gov.cn/srcsite/A26/s8001/201112/t20111228_167340.html. Accessed 30 Sept 2020
5. Gao, R.X., Hu, J.Y., Yang, J., Fan, Y.Q., Mo, L.: Study on the main problems and countermeasures of campus reading promotion for primary and secondary school students, using the primary schools in Yuexiu District in Guangzhou as subjects. Psychol. Explor. **39**(02), 140–145 (2019)
6. Guangzhou Intelligent Reading Project Group: White Paper on Reading for Primary School Students in Guangzhou (2019)
7. Mo, L., Liu, X.L., Zheng, H.Y.: The study on the main problems and countermeasures of campus reading in primary and secondary schools. Minist. Educ. Brief. (Univ. Column) **118**(01), 1–6 (2017)
8. Guangzhou Government. http://www.gz.gov.cn/zwgk/zjgb/zfgzbg/content/mpost_2840352.html. Accessed 30 Sept 2020
9. Gao, R.X., et al.: An empirical research based on word association tests: the development of learning concepts in chinese students from the perspective of "virtue and mind model". J. East China Normal Univ. (Educ. Sci.) **38**(04), 83–95 (2020)
10. Chinese Social Sciences Net. http://www.gz.gov.cn/zwgk/zjgb/zfgzbg/content/mpost_2840352.html. Accessed 30 Sept 2020

A Study on College Students' Self-regulated Online Learning in the Home Study Context

Wenyi Chen(✉)

Zhongkai University of Agriculture and Engineering, No. 501, Zhongkai Road, Guangzhou, Guangdong Province, China
winniecwy@126.com

Abstract. Under the circumstance of school closure due to the outbreak of COVID-19, one of the greatest challenges for the sudden transition from traditional face-to-face teaching to fully online learning lies in students' inadequate capability in self-regulated learning (SRL). This study investigates how the first-year undergraduate students regulated their online learning in an EFL Small Private Online Course (SPOC) during the pandemic period. A revised version of Jansen et al.'s (2018) Self-regulated Online Learning Questionnaire (SOL-Q) is applied to collect information of the participants' SRL ability. Results indicated that the participants' overall self-regulated online learning ability was generally at an intermediate level, however there's still room for improvement. Specifically, the participants demonstrated limited capability in respect of meta-cognitive skills in the preparatory phase and persistence. A positive correlation was found between metacognitive skills and time management. Teaching implications are provided on how to improve students' metacognitive skills, time management and persistence in future online instructions.

Keywords: Online learning · Self-regulated learning · SPOC · COVID-19

1 Introduction

The COVID-19 pandemic has brought a profound influence on people's life and education is non-exception. On January 19th, Chinese Ministry of Education launched an initiative entitled "Disrupted Classes, Undisrupted Learning" (Huang et al. 2020).

China is the first country to launch such a policy. With more than 278 million students shifted to online learning, technology has made it possible to mitigate the crisis of COVID-19 and provide "undisrupted learning" (UNESCO IITE, 2020). Massive Open Online Courses (MOOCs) and other forms of online learning such as Small Private Online Courses (SPOCs), once a supplement to traditional classroom teaching quickly shifts into a replacement for it overnight. Due to the autonomy of students in this type of education, students are required to regulate their learning to a greater extent than students in traditional, face-to-face education.

L.-K. Lee et al. (Eds.): ICTE 2020, CCIS 1302, pp. 189–200, 2020.
https://doi.org/10.1007/978-981-33-4594-2_16

1.1 SRL in MOOCs

SRL refers to students' systematic effort to manage their learning process to achieve goals (Pintrich 2004; Zimmerman and Schunk 2011). Research indicates that self-regulation is of critical importance in determining students' successful learning experiences in an online learning environment (Cho and Kim 2013). Self-regulated learners effectively manage their time and learning resources. They set goals, plan ahead, and consistently monitor and reflect on their own learning process (Pintrich 2004; Zimmerman 2011).

Massive Open Online Courses are courses offered by universities on designated MOOC platforms (e.g., edX and Coursera). They are a particular form of online education in which a great amount of autonomy are provided (Hew and Cheung 2014; Kizilcec and Halawa 2015). SRL becomes of greater importance for learner success when the learning process is less externally regulated by the teacher (Jansen et al. 2018). Learners must then manage their learning to a greater extent, making SRL more critical (Beishuizen and Steffens 2011; Wang et al. 2013). The necessity of SRL in MOOCs in combination with the increased number of MOOCs offered (Allen and Seaman 2016) has made researching learners' SRL in this context valuable. Initially, research on SRL in MOOCs and other forms of online education made use of questionnaires, showing positive correlations between self-reported SRL activity and course completion (Wang et al. 2013; Yukselturk and Bulut 2007).

1.2 SPOC

Small Private Online Courses (Fox 2013) are small-scale online courses and represent a specific, defined form of fully online education. Fillius et al. (2018) argue that Small Private Online Courses (SPOCs) may be a type of online learning for higher education that has good potential to promote deep learning. SPOCs have a fixed start- and end date, with an instructor scaffolds the students through the learning process. SPOC do not emphasize the various and complete of resource, but emphasize on personalized characteristics of resources (Armando and David 2014).

SPOC is a new hybrid mode of the integration of online learning and traditional classroom for small scale, specific population by using MOOC resources and online evaluation, communication and other functions (Teplechuk 2013). Since 2013, the domestic colleges and universities have conducted about SPOC teaching attempt. A total of 85 SPOC/MOOC courses from 51 universities are on line on Chinese MOOC platform (Wang 2016). The relevant SPOC construction has been the bridge of MOOC lesson from the "platform technology" to "classroom teaching" (Bulfin et al. 2014). However, according to online SPOC courses and literature on SPOC, most researches and constructions of SPOC still stay at the stage of providing curriculum resources for learners (Caswell et al. 2008), and lack of monitoring and learning evaluation methods (Downes 2013). Filius et al. (2018) found that during the delivery phase of the SPOC, instructors found it difficult to monitor the progress of their students and find out what students needed.

1.3 Current Study

Little empirical research has been conducted in SPOC with regard to SRL. Learners however often struggle to successfully regulate their learning process. Due to the autonomy provided to learners in SPOCs, learners are required to regulate their learning to a greater extent than students in traditional, face to face education. It is therefore important to measure learners' SRL in SPOCs, especially during the pandemic period when courses at all levels are completely given online. This study presents a SRL investigation implemented in a SPOC *College English II* to measure learners' SRL in the home study context during the pandemic period.

2 Research Design

2.1 Research Questions

Based on the research purpose, three research questions were formulated:

1. What is the status quo of students' overall self-regulated online learning ability?
2. What are students' self-regulated online learning ability in terms of the seven sub-scales?
3. Are there correlation between students' abilities in the seven sub-scales?

2.2 Participants

The teaching experiment with the questionnaire survey took place in March of 2020. Four classes of undergraduate students from Zhongkai University of Agriculture and Engineering (ZKU) were randomly chosen to participate in the present study (245 students in total). They were first-year non-English majors who had just passed the National College Entrance Examination. They had taken the course of *College English I* in the previous semester and were already familiar with the teaching style of their English teacher.

2.3 Research Instrument

The research instrument in the present study is a questionnaire adapted from the Self-regulated Online Learning Questionnaire – Revised (SOL-Q-R; Jansen et al. 2018). As is initially developed for fully online courses with a focus on individual learning act ivities, they argue that it is transferable for similar settings (Jansen et al. 2016). The questionnaire is originally consisted of 42 items divided over 7 scales: metacognitive activities before learning (7 items, $\alpha = 0.87$), metacognitive activities during learning (7 items, $\alpha = 0.80$), metacognitive activities after learning (6 items, $\alpha = 0.85$), time management (5 items, $\alpha = 0.69$), environmental structuring (4 items, $\alpha = 0.81$), persistence (7 items, $\alpha = 0.86$), and help seeking (6 items, $\alpha = 0.92$). It was adapted and translated into Chinese by the present author into 30 items which also includes 7 scales of students' meta-cognitive activities before learning, students' meta-cognitive activities during learning, students' meta-cognitive activities after learning, time management, environmental structuring, persistence and help seeking. The rationale behind this adaptation is that since the SOL-Q-R was developed by Jansen et al. for the evaluation of the learners' self-regulated online learning it is suitable for this similar setting, however in order to help the EFL learners to get a better understanding and to achieve better validation, the present authors translated it into Chinese before the investigation. During the translation, there was slight adaptations, some items with similar function were combined together, some were deleted, wording were adjusted and finally the present questionnaire were consisted of 30 items (see Table 1).

Table 1. Questionnaire adapted from SOL-Q-R

SOL-Q scales	Phases	No. of items	Items
Meta-cognitive skills	Before learning	4	1, 2, 3, 4
	During learning	6	5, 6, 7, 8, 9, 10
	After learning	5	26, 27, 28, 29, 30
Time management		3	11, 12, 13
Environmental structuring		4	14, 15, 16, 17
Persistence		4	18, 19, 20, 21
Help seeking		4	22, 23, 24, 25

Students were instructed to respond to the items on a 5-point Likert scale (1 = strongly disagree to 5 = strongly agree) in terms of their behavior in the SPOC *College English II*. As is shown above in Table One, the whole questionnaire includes seven sub-scales, three phases and 30 items. they are the scale of meta-cognitive skills before learning (4 items),the scale of meta-cognitive skills during learning (6 items), the scale of meta-cognitive skills after learning (5 items). The above three sub-scales also refer to the three phases in the meta-cognitive skills respectively. Besides, there are the scale of time management including 3 items, the scale of environmental structuring 4 items, the scale of persistence 4 items and the scale of help seeking 4 items.

2.4 Procedure

The teaching materials of the present SPOC was consisted of two parts. The first part, from Weeks 1 to Week 6, which amounted to 30%, was a SPOC borrowed from an open online course *Internet and College English* provided on one of the leading MOOC platforms in China——iCourse (https://www.icourse163.org/). It included five modules of technology-enhanced listening, speaking, reading, writing and translating. The teaching materials of Week 7 to Week 16, *College English II,* which amounted to 70% was provided by the ZKU's teachers on the platform Chaoxing. The teaching content of the present study is shown below (Fig. 1).

In Week 4, a live class on oral English was synchronized which meant all students of the same class attended the online classroom at the same timeslot and had real-time

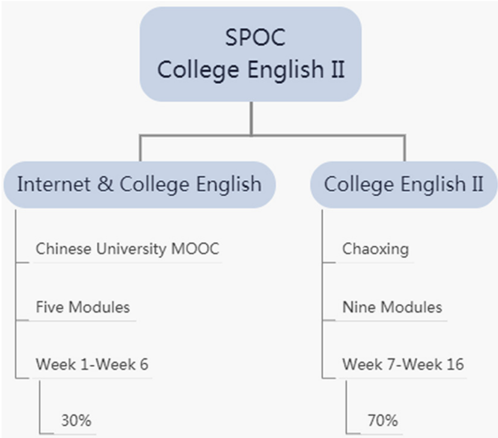


Fig. 1. Teaching material of the present SPOC *College English II*

interactions. During the class, the teacher, TA and the students met online and practised oral English by sharing opinions on certain topics prepared by the teacher. After the online class, the students continued to make improvement on their oral English homework and oral task of recording a dialogue with one or more of his classmates under the instruction given to them during the online class. There were homework assignments weekly in accordance. The students should finish the homework and upload the documents to the e-learning platform (ChaoXing) .

The questionnaire was added as a voluntary activity at the end of Module 2, which was exactly at the end of the live class in week 4. Learners were invited to fill out the questionnaire as a voluntary activity within the learning environment. Among the 245 students who participated in the live class, a total of 141 learners filled out the SOL-Q, in terms of their behavior in the specific SPOC *College English II* class. Responses of 141 students were finally used for analyses (Mean age = 19, 80% male). In this case study, the materials and research data were obtained from the class in March 2020.

2.5 Data Collection and Analyses

Online questionnaire tool Wenjuan Wang (https://www.wenjuan.com) was adopted to deliver the present questionnaire and collect data back. One sample T-test (a chosen value of 3.75 is determined, since it's a 1–5 scale questionnaire, the researcher regard as upper intermediate as above 75% and is therefore 3.75) was conducted to analyze the students' SRL in terms of the 7 scales, and comparison were made between each scale in SOL-Q-R. Furthermore, a Pearson correlation study was conducted to analyze the correlation between the seven scales of the self-regulated learning in the SPOC. M-score, T-score and Pearson correlation analysis were used as statistics.

3 Results and Discussion

The results of the analyses indicate that the learners' overall SRL is generally at an intermediate level, but there is still room for improvement for some specific scales. Meta-cognitive skills during the learning process and environmental structuring in the SOL-Q-R are generally at an upper intermediate level. However, the scores of scale meta-cognitive skills before the learning process, time management, persistence and help seeking witness a comparatively lower level. A positive correlation was found between metacognitive skills and time management.

3.1 Overall Performance of Students' SOL

Table 2 presents the mean score and the one sample T test results of learners' SOL in the SPOC (see Table 2).

Table 2. One sample T-test results of learners' SRL in the SPOC

SOLQ scales	M	SD	P (two tailed)
Metacognitive skills	3.97	0.81	3.74161E−10**
Activities before	3.86	0.80	0.180
Activities during	3.98	0.77	1.7538E−05**
Activities after	3.96	0.85	0.007*
Time management	4.01	0.79	0.079
Environmental structuring	4.17	0.81	0.003*
Persistence	3.87	0.82	0.100
Help seeking	3.99	0.87	0.054

Note: The mark "" indicates significant difference, while "**" indicates very significant difference.*

The results of SRL analyses indicate that the learners' overall SRL is generally above an intermediate level (mean score = 3.97, SD = .81, $p < .01$), which indicates that learners' SRL measured in this SPOC is significantly above the intermediate level. A closer observation on the three sub-scale of meta-cognitive skills, metacognitive activity before, during and after learning indicates that learners report significantly higher level of SRL during (mean score = 3.98, SD = .80, $p < .01$) and after learning (mean score = 3.96, SD = .80, $p < .05$), the scale environmental structuring (mean score = 4.17, SD = .81, $p < .05$) also show an upper intermediate level. However, students differ in their abilities to accurately regulate their learning (e.g., Azevedo and Cromley 2004). There is still room for improvement for most of the scales, which are not significantly higher than an intermediate level. The scales metacognitive activity before learning, the scale time management, the scale persistence and the scale help seeking are not significantly higher than an intermediate level, which indicates that there is still room for improvement.

3.2 Students' Abilities in the Seven Sub-scales

Table 3 presents items that rank the highest in the result of learners' SRL in the SPOC .

Table 3. Top 5 items that ranks the highest in the result of learners' SRL in the SPOC

Scales	Items	M	SD
ES	15. I find a comfortable place to study for this online course	4.29	0.72
ES	16. I have a regular place set aside for studying for this online course	4.23	0.89
P	12. Even when materials in this online course are dull and uninteresting, I manage to keep working until I finish	4.17	0.76
ES	14. I choose the location where I study for this online course to avoid too much distraction	4.13	0.78
MSD	9. I ask myself questions about how well I am doing while learning something in this online course	4.03	0.74

Note: In Column 1 the scale "ES" stands for Environmental structuring, "P" stands for Persistence, "MSD" stands for meta-cognitive skills during learning.

As is shown in the above table, items at the top list are mostly concerning with the scale of environmental structuring, which amounts to 60%, with three out of the total five items. Among the tops five items, learners apparently display better management in their environmental structuring, with Item 15 "I find a comfortable place to study for this online course." rank the highest, Item 16 "I have a regular place set aside for studying for this online course." ranks the second, and Item 14 "I choose the location where I study for this online course to avoid too much distraction." being the fourth. According to Williams and Hellman (2004), the openness in time and place makes students solely responsible for their time and environment management. This may indicate that learners stick great importance to a suitable environment for online learning, while at the same time it also demonstrate a truth that with the development of China's economy, today's youngsters in China put more emphasis on the suitable environment for learning in the home study context.

The persistence factor "Even when materials in this online course are dull and uninteresting, I manage to keep working until I finish." ranks the third in the top five list, which reports learners' effort to persist on finishing the learning tasks regardless of its dullness. There are views that, SPOC = Classroom + MOOC (Armando and Berkeley 2013). In the home study context, SPOC is actually a compulsory course on campus for the students and specific learning goals and requirement are given by the instructors at the beginning of the course. In order to pass the course and earn the credit, strict regulations are needed. Finally, the meta-cognitive skill in the performance phase "I ask myself questions about how well I am doing while learning something in this online course." is listed on the fifth of the list, this displays learners' learning strategy of self-reflection during the learning process. This may be explained by the learners' self-awareness in the learning process, which in another way may also indicate that learners are motivated to learn the course.

Table 4. Top 5 items that ranks the lowest in the Result of learners' SRL in the SPOC

Scales	Items	M	SD
MSB	4. I set short-term (daily or weekly) goals as well as long-term goals (monthly or for the whole online course)	3.7	0.87
P	19. When I am feeling bored studying for this online course, I force myself to pay attention	3.72	0.89
HS	25. When I do not fully understand something, I ask other course members in this online course for ideas	3.77	0.97
MSA	26. I ask myself how well I accomplished my goals once I'm finished working on this online course	3.82	0.9
MSB	3. I set specific goals before I begin a task in this online course	3.82	0.77

Note: In Column 1 the scale "MSB" stands for Meta-cognitive skills before learning, "P" stands for Persistence , "HS"stands for help seeking, "MSA" stands for Meta-cognitive skills after learning.

Table 4 presents items that rank the lowest in the result of learners' SRL in the SPOC.

As is shown in the above table, items with the lowest score are much more diverse. Among the five items, there are totally four scales on concern, with two items concerning meta-cognitive skills (activities before learning), one with meta-cognitive skills (activities after learning), persistence and help seeking respectively.

Meta-cognitive skills in the preparatory phase witnesses the lowest in the list. This skill concerns goal setting and task definition, which are often carried out by the lecturer in traditional education (Jansen et al. 2016). The lack of such skill may indicate that learners are still not used to the online learning model. Owing to the sudden outbreak of the epidemic, students are confined to have online course at home, however, most of them haven't got much online learning experience and haven't received much training to adapt themselves to the online learning setting. Persistence factor "When I am feeling bored studying for this online course, I force myself to pay attention." ranks the second lowest. This item reports students' absent-mindedness during the learning process, which may indicate that without specific learning goal learners' persistence may be at a lower level. The scale help seeking "When I do not fully understand something, I ask other course members in this online course for ideas." ranks the fourth lowest. That the Chinese students are quiet and rarely raise questions is a common phenomenon in traditional classroom which are usually labeled as "being shy", however the lack of motivation to know the truth behind this phenomenon cannot be ignored. Meta-cognitive skill during learning "I ask myself how well I accomplished my goals once I'm finished working on this online course," ranks the fifth lowest. According to Zimmerman (2002), self-regulating students reflect on their performance by comparing their achievements to the goal they set. This indicates the close relation between the two meta-cognitive skills before learning and after learning. Learners in the present study are not good at setting either long-term goals or short-term goals at the beginning of the course, which may result in the lack of reflection after learning.

3.3 Correlation Between the Seven Sub-scales

Table 5 presents the results of a Pearson correlation study between the seven scales of the self-regulated learning in online learning based on the survey result of the questionnaire.

Table 5. Correlation between the seven sub-scales of SRL measured.

Scale	Before	During	After	Time	Environment	Persistence	Help
Before	1	0.800	0.734	0.803	0.658	0.735	0.671
During		1	0.782	0.854	0.708	0.746	0.662
After			1	0.812	0.751	0.730	0.777
Time				1	0.710	0.737	0.712
Environment					1	0.627	0.683
Persistence						1	0.638
Help							1

Correlations were analyzed between each two scales of SRL on learners' online learning performance. As is shown in the above table, correlation coefficient values is above .620 between each two scales among the seven scales of SRL, which is in alignment with previous research, "Self-regulated learners are actively involved in their own learning process, not only during learning (performance phase), but also before (preparatory phase), and after learning (appraisal phase)" (Jansen et al. 2018, p. 116). Correlation between metacognitive skills before learning and metacognitive skills during learning witnesses quite high positive correlation ($r = .80$), which may indicate that the skills of goal setting and course planning in the preparatory phase will influence the students' learning performance. This finding is also in alignment with the previous findings that self-regulating students reflect on their performance by comparing their achievements to the goal they set and based on this evaluation, students adapt their study strategies in the-sometimes very near-future (Zimmerman 2002; Pintrich 2000; Winne and Hadwin 1998). Another high positive correlation lies between the scale time management and metacognitive skills at all phases. The correlation coefficient values between them are all above .80, which indicates very strong correlations between the scale time management and the scale metacognitive skills before learning, during learning and after learning. The scale time management is therefore a very important factor for self-regulated learners who are actively involved in their own learning process.

4 Conclusion and Implications

As Allen and Seaman (2016) pointed out that the necessity of SRL in MOOCs in combination with the increased number of MOOCs offered has made researching learners' SRL topical and valuable. This is even more true for SPOC in the home study context. The present study investigates how learners self-regulated their online learning in an EFL

SPOC during the pandemic period. After the outbreak of COVID-19, one of the greatest challenges on campus students are suddenly forced to face lies in students' inadequate capability in self-regulated learning (SRL). Four classes of first year undergraduate students were randomly chosen to participate in the present study. Results indicated that the participants' overall self-regulated online learning ability is generally at an intermediate level, however there's still room for improvement. Specifically, the participants demonstrate limited capability in respect of meta-cognitive skills in the preparatory phase and persistence. This requires additional training on goal setting and course planning of students enrolled in SPOC compared to students in traditional education. A positive correlation is found between metacognitive skills in all phases and time management. It is suggested that future online instruction focus more on a much more detailed direction with guidance step by step on improving students' time management, metacognitive skills in the preparatory phase, the performance phase and the appraisal phase. The positive correlation between each two scales also indicates that improvement on any of the scales will bring positive effect on the other scales. Instructors should therefore strive to help learners improve their SRL in SPOC in the home study context.

5 Limitation

It's a little bit pity that the present study only collected 141 questionnaires out of the 141/245 participants that filled in questionnaires, probably owing to the fact that it is a voluntary activity at the end of the live class. In future research, the researcher would try to collect data in a more effective way, for instance, by giving bonus to students who fill in the questionnaire.

Acknowledgement. My heartfelt thanks go to the four anonymous reviewers and Professor Chen Zexuan from Southern Medical University for their invaluable feedbacks on earlier versions of this paper.

References

Allen, I.E., Seaman, J.: Online report card: tracking online education in the United States (2016). https://files.eric.ed.gov/fulltext/ED572777.pdf

Azevedo, R., Cromley, J.G.: Does training on self-regulated learning facilitate students' learning with hypermedia? J. Educ. Psychol. **96**, 523–535 (2004)

Beishuizen, J., Steffens, K.: A conceptual framework for research on self-regulated learning. In: Carneiro, R., Lefrere, P., Steffens, K., Underwood, J. (eds.) Self-Regulated Learning in Technology Enhanced Learning Environments, pp. 3–19. Sense Publishers, Rotterdam (2011)

Bulfin, S., Pangrazio, L., Selwyn, N.: Making 'MOOCs': the construction of a new digital higher education within news media discourse. Int. Rev. Res. Open Distrib. Learn. **15**(5), 290–305 (2014)

Caswell, T., Henson, S., Jensen, M., Wiley, D.: Open content and open educational resources: Enabling universal education. Int. Rev. Res. Open Distance Learn. **9**, 1–11 (2008)

Cho, M.-H., Kim, B.J.: Students' self-regulation for interaction with others in online learning environments. Internet High. Educ. **17**, 69–75 (2013)

Downes, S.: MOOCs will ultimately play a transformational role. In: Half an Hour, Blog (2013). http://halfanhour.blogspot.co.uk/2013/11/moocs-will-ultimately-play.html. Accessed 6 Nov 2013

Filius, R.M., de Kleijn, R.A.M., Uijl, S.G., Prins, F.J., van Rijen, H.V.M., Grobbee, D.E.: Challenges concerning deep learning in SPOCs. Int. J. Technol. Enhanced Learn. **10**(2), 111–127 (2018)

Fox, A.: From MOOCs to SPOCs. Commun. ACM **56**(12), 38–40 (2013)

Hew, K.F., Cheung, W.S.: Students' and instructors' use of massive open online courses (MOOCs): motivations and challenges. Educ. Res. Rev. **12**, 45–58 (2014)

Huang, R.H., Liu, D.J., Tlili, A., Yang, J.F., Wang, H.H., et al.: Handbook on facilitating flexible learning during educational disruption: the Chinese experience in maintaining undisrupted learning in COVID-19 outbreak. Smart Learning Institute of Beijing Normal University, Beijing (2020)

Jansen, R.S., van Leeuwen, A., Janssen, J., Kester, L., Kalz, M.: Validation of the self-regulated online learning questionnaire. J. Comput. High. Educ. **29**(1), 6–27 (2016). https://doi.org/10.1007/s12528-016-9125-x

Jansen, R.S., Van Leeuwen, A., Janssen, J., Kester, L.: Validation of the revised self-regulated online learning questionnaire. Lifelong Technol. Enhanced Learn. **11082**, 116–121 (2018). https://doi.org/10.1007/978-3-319-98572-5_9

Kizilcec, R.F., Halawa, S.: Attrition and achievement gaps in online learning. Presented at the L@S, pp. 57–66. ACM Press, Vancouver (2015)

Pintrich, P.R.: The role of goal orientation in self-regulated learning. In: Boekaerts, M., Pintrich, P.R., Zeidner, M. (eds.) Handbook of Self-Regulation, pp. 451–502. Academic Press, San Diego (2000)

Pintrich, P.R.: A conceptual framework for assessing motivation and self-regulated learning in college students. Educ. Psychol. Rev. **16**(4), 385–407 (2004). https://doi.org/10.1007/s10648-004-0006-x

Puustinen, M., Pulkkinen, L.: Models of self-regulated learning: a review. Scand. J. Educ. Res. **45**, 269–286 (2001)

Institute for Information Technologies in Education, UNESCO. Online and Open Education in Shanghai: Emergency Response and Innovative Practice during COVID-19 Pandemic (2020). https://iite.unesco.org/wp-content/uploads/2020/06/Online-and-Open-Education-in-Shanghai-Emergency-Response-and-Innovative-Practice-during-COVID-19-Pandemic.pdf

Teplechuk, E.: Emergent models of massive open online courses: an exploration of sustainable practices for MOOC institutions in the context of the launch of MOOCs at the University of Edinburgh. MBA Dissertation, University of Edinburgh (2013)

https://www.era.lib.ed.ac.uk/bitstream/1842/7536/1/MOOCs_MBADissertationTeplechuk_Master.pdf

Wang, C.-H., Shannon, D.M., Ross, M.E.: Students' characteristics, self-regulated learning, technology self-effcacy, and course outcomes in online learning. Distance Educ. **34**, 302–323 (2013)

Wang, X.H., Wang, J.P., Wen, F.J., Wang, J., Tao, J.Q.: Exploration and practice of blended teaching model based flipped classroom and SPOC in higher university. J. Educ. Pract. **7**(10), 99–104 (2016)

Williams, P.E., Hellman, C.M.: Differences in self-regulation for online learning between firstand second-generation college students. Res. High. Educ. **45**(1), 71–82 (2004). https://doi.org/10.1023/B:RIHE.0000010047.46814.78

Winne, P.H., Hadwin, A.F.: Studying as self-regulated learning. In: Hacker, D.J., Dunlosky, J., Graesser, A. (eds.) Metacognition in Educational Theory and Practice, pp. 277–304. Lawrence Erlbaum Associates, Mahwah (1998)

Yukselturk, E., Bulut, S.: Predictors for student success in an online course. Educ. Technol. Soc. **10**(2), 71–83 (2007)

Zimmerman, B.J.: Becoming a self-regulated learner: an overview. Theory Pract. **41**, 64–70 (2002)

Zimmerman, B.J., Schunk, D.H.: Handbook of Self-Regulation of Learning and Performance. Routledge, New York (2011)

Methods and Techniques to Improve International Learning Experiences Online

Wenbin Ji(✉), Pan Li, Meihua Wang, and Jinbin Zhang

China Meteorological Administration Training Centre, Beijing 100081, China
jiwenbin@cma.cn

Abstract. International courses in physical classroom tend to be infeasible under the influence of epidemic diseases, such as the COVID-19. Consequently, the transition of teaching activities to online events, which is advantageous in reducing transportation and other expenses, has been acknowledged as an effective and practical solution. However, it also brings new challenges to the teaching in quarantine for vocational education and training, because most of learners at their working places may need to do other things in the online environment during their online class time, such as temporary works, attending meetings and receiving instant messages. In addition, they are usually from different cultures and time zones. Therefore, it is reasonable to create online engaging learning experiences with some careful planning. Taking the International Distance Training Course on Short-term Climate Monitoring and Prediction in Disaster Prevention and Mitigation (from 17th to 31st, May, 2020) as an example, this paper introduces some relevant practical instructional designs, collaborative methods and interactive techniques for improving learning experiences online, and analyses their effects.

Keywords: Online learning · Collaborative tools · Student engagement · Vocational training · Instructional design

1 Introduction

Meteorology and Hydrology are usually recognized as disciplines within the global scope, and their related works depend majorly on worldwide cooperation and collaboration. The interest among meteorologists and hydrologists from the National Meteorological and Hydrological Services (NMHSs) in attending international training courses on atmospheric science is increasing steady. And their attendances are aiming to strengthen expertise and communicate experience especially on the global data collection, application and sharing. China Meteorological Administration Training Centre (CMATC) has been conducting a range of international courses face-to-face for trainees as the Beijing Component of WMO Regional Training Centre (RTC) and one of the WMO/CGMS Virtual Laboratory Centers of Excellence (CoE). However, those courses are prone to be strongly affected by some unexpected factors, such as transportation budget, traffic delay, natural disaster, and epidemic disease. As a result, their planning flexibility is

L.-K. Lee et al. (Eds.): ICTE 2020, CCIS 1302, pp. 201–212, 2020.
https://doi.org/10.1007/978-981-33-4594-2_17

sometimes relatively insufficient, especially in the case of established schedule change during the execution process.

According to Parrish (2008), online learning has been an important force in helping the meteorological community share knowledge and experience while facing the new challenges and opportunities of instructional innovations. Spangler (2008) holds an opinion that there is no single way to best educate both emergency professionals and the general public, but distance learning is one technique that has proven successful in reaching large numbers of people. Kelsch and Laing (2011) emphasize that, web-based modules are used extensively worldwide by students and instructors in the university community, as well as professionals and enthusiasts in both government and private enterprise. Therefore, a series of online international courses with carefully planning have been released in succession by CMATC from the beginning of 2020, based on the experiences and lessons learnt from former experimental attempts.

1.1 Background

The primary role of universities is to develop background knowledge and critical thinking skills rather than prepare students for specific job tasks, and some universities provide students with knowledge of meteorology and/or hydrology [1]. It is the services provided by all the NMHS training departments and third parties that ensure trainees are well equipped with expertise and competencies. However, the contradiction between work and study is usually an inevitable issue that vocational training suffers most. In addition to the threshold of professional specialty and the need to comprehend abstract and complex theories, mastering weather related operational technologies and methods in a shortened term brings the education and training practitioners more challenging works.

1.2 Transitional Trend

Taking the meteorological education and training events sponsored by the Chinese meteorological departments as examples, online meteorological courses have been designed and carried out initially in the form of audio/video broadcasting over Internet/Intranet in domestic scope. On the other hand, international training courses are generally designed as face-to-face events and have been executed for decades. In order to distinguish the differences of the online international course and the previous courses, domestic online course and international face-to-face course are considered as "traditional courses" within this paper.

Recently, team building for professionals and capacity building in education and training become the priorities of meteorological tasks, which benefits from the modernization development proposal to boost meteorological services. As a result, the condition for conducting international online education and training is ready. In order to carry out online international course normally, some experimental works started in 2018, and the "distance learning course on the maintenance of meteorological observation instruments" was launched firstly. That course unexpectedly gained far more trainees, nearly 10 times, than common face-to-face course at the same period, for the advantages in recruiting trainees and overcoming geographical barriers.

However, the traditional ways of training related to physically attending classroom, including international face-to-face courses, are infeasible due to the issues like travel security and high cost, facing the unexpected global epidemic disease, especially the COVID-19. As a result, most of the courses trend to be transplanted online. However, some feedbacks from learners complain about the reduction of training quality, in terms of insufficient interactivities, network disturbing and concentration distracting. Therefore, there should be some improvement to create engaging learning experiences online with careful design, based on the previous international course practices.

1.3 New Challenges

More challenges for improving international online learning experiences need to be dealt with, which are: 1) Located in different time zones, learners keep independent schedules and routines; 2) Learners are from different cultures and backgrounds, hence there should be relatively high capability requirements for cross cultural communication; 3) Learners from Asia, Africa and Latin America used to report their limitations in local network bandwidth, and intermittent online multimedia resources or file submission. In addition, the situation that learners may be at work when some online education and training events are conducted cannot be ignored.

2 Instructional Designs to Hatch Engagement

The transition to online course is a comprehensive reconstruction process from the present state, and methods like pushing all the face-to-face components online while keeping the whole structure intact sometimes yield insignificantly. The aim of instructional design is to identify optimal solution for its strategy and training needs, solve or avoid problems and arrange suitable resources in a specific situation. As Parrish and Spangler (2008) propose educators and trainers to focus on the essence of learning, not just technology, instructional designing benefits the whole course.

2.1 Triggered by Participants' Demands

Course engagement usually determines its learning experience, and it varies according to some learners with different backgrounds and interests. Pre-training questionnaire is recognized as an effective method in reflecting whether and to what extent the training program satisfies the demand of learners. And it also provides a clue for instructors to know the general condition and level of students before courses start. Learners tend to show greater interest and report better learning experiences and outcomes, if the courses are tailored according to their own levels. Moreover, learners can shorten the process of evoking past memories and knowledge that are kept in long term storage, and quickly get into their learning states with the aids of questionnaires.

The questions in pre-training questionnaires involve: The highest academic degree you are awarded? How many years have you worked in your current organization? Have you ever attended this or other organizations' training courses before? Have you ever attended training programs you consider very useful for your current career? Have you ever attended similar trainings before? What subjects would be organized in the future? What do you consider interesting for yourself or your colleagues in your organization that you would like to attend? What is the most favorite topic for you?

Besides, post-training questionnaires can also provide supplementary information to determine the needs for further instructional design. According to the "learning cycle" [1], the progression of assessing and evaluating the learning process stage represents the current needs and outcomes, and it can also assist to identify the learning needs of next stages. So, the collection and analysis of demand help improve the learning experiences in a cycle.

2.2 Using the Engagement Curves

Learners have different thoughts and feelings when they first become engaged, when the pattern of the instruction becomes evident and accepted (or resisted), and when learning is approaching its culmination [2]. Parrish (2008) points out that narrative diagrams can help plan for a learning experience which will have necessary structure to develop deep engagement and a fulfilling outcome by plotting the sequence of topics and learning activities. In this case, the diagram created can be called the engagement curve, depicting the story during the learner's journey into the pending knowledge.

Figure 1 illustrates the narrative diagram in preparing this online course. The Y-axis in the diagram is used for fictional narratives, and it typically describes the level of action or complication. From the perspective of engagement curve, the course was designed to be decomposed into 11 activities, based on the social constructionist education conception. And every module in the middle stage was used to frequently stimulate the initial engagement and try to avoid any potential experience consumption.

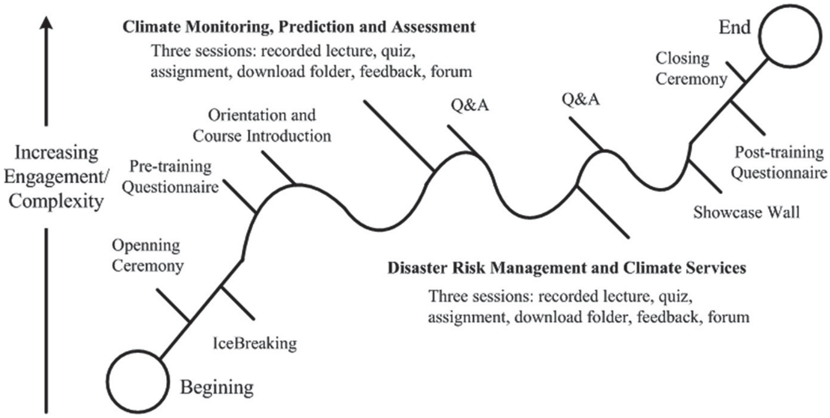


Fig. 1. The engagement curve in preparing the online course.

2.3 Framework Construction and Knowledge Sharing

A well planned online international course indicates that its knowledge framework accords with the cognitive law of learners, which means that learners may follow a set of directed sequence or mapped route for navigation. And instructional designers work on embedding the learning resources and modules to their right contexts to make sure confusion clearing, otherwise learners may feel difficult to follow the schedule and reduce their enthusiasms and finally the learning experiences.

Disaster risk reduction is noted as one of the priority areas of WMO strategic objectives, which are considered to determine the long-term success of the WMO Global Campus initiative. And this course works as a part of the meteorological education and training communities, and contributes to the WMO Global Campus in resource sharing. The course aim is to enhance the trainees' understanding in meteorological disaster prevention and mitigation, provide the trainees with related operational approaches on short-term climate prediction, and enhance the technical exchange and future cooperation in the meteorological field. After it, participants would be able to build up technical knowledge on short-term monitoring and prediction approaches, and get a better understanding on meteorological disaster prevention and mitigation.

2.4 Collaboration Network Identification and Utilization

In the context of online international training, teamwork collaboration can be viewed as a virtual network based on the relationships among participants and events, and all mutual actions online such as tutor support, peer support, comment and reply are also regarded as links of relationships.

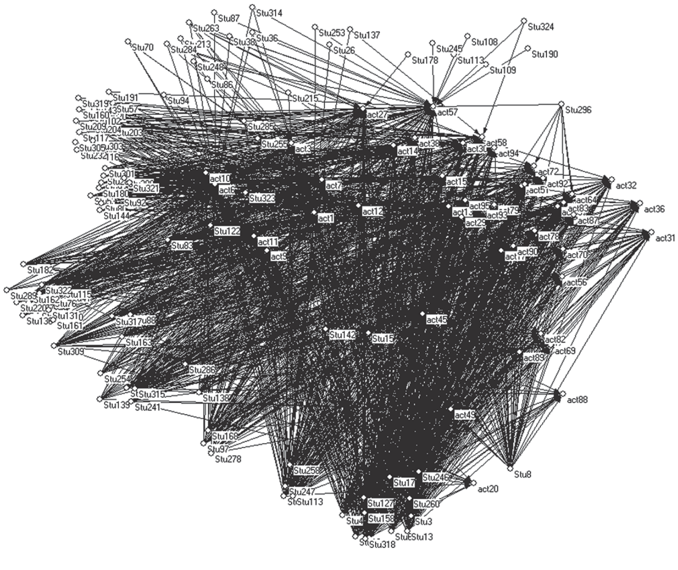


Fig. 2. The teamwork collaboration network map in online course.

Figure 2 shows a map of collaboration based on the data from this course using one of the Pivot MDS [8] algorithms. The number of vertices is 419, and the total number of links is 13,856. The average degree is one of the critical factors to measure the network tightness and scale, which is representative of engagement and even learning experience, and in this case it is 66.14. Although the ideal state of network is full connection, the collaboration network is relatively sparse, and its density is just 0.0789 in this case, which is far less than 1. As a matter of fact, the distance of this network is much shorter than it looks, after adding enough distant collaboration, which means the average distance from learners to teaching modules and other participants will be shorter, and learners may feel easier to find supports. However, building collaboration over long distance is considered to be more challenging, which needs adequate instructional interactions before establishing effective communication.

In order to make the collaborations more engaging in the global scale, building collaborations locally and over long distances are proposed. The local collaborations resemble learning communities, and learners will benefit from local supports in sharing knowledge and experiences while facing the new challenges and opportunities of instructional innovations. In the global scale, nodes that can build long distance collaboration should be identified and assigned the roles of focal points or facilitators, including the working group members, during the preparation and implementation of the online events.

3 Techniques and Tools to Improve Experiences

Today nearly all computers have multimedia capabilities, and the World Wide Web not only rivals the classroom as a common venue for training and education, but has also become the primary playground and communications tool for millions of people around the world. For these reasons, the emphasis on instructional methodology has grown, and a higher percentage of sessions address innovative uses existing technologies [7]. There are emerging systems and tools ready for customization, so the point is that how to integrate them as an organism, and guarantee the smoothly functioning as well as flexible and robust.

3.1 Interactive Training System

Learning Management System (LMS) works as the hub for training programs. In order to engagingly conduct the international online course, Moodle [9] has been adopted in its "subject" format to follow the previously agreed engagement curve. In the course, most of the modules were asynchronous to overcome the time differences. The course was designed into 3 parts, the beginning, middle and end.

In the first part of the course, an ice-breaking session was built as a pre-course collaboration, and its feedbacks showed that participants could benefit from quickly overcoming the feeling of isolation, getting familiar with peers and team building. Besides, opening ceremony, schedule description, course introduction and pre-training questionnaire were established. All the sessions above together served as the beginning part of the course, which drew enough attentions from learners, prepared students to start the voyage along the designated route and developed trust that the tension can be resolved.

In the middle of the course, 2 different subjects worked successively as the core body and had their own instant chat rooms and forums, which facilitated learners to communicate with instructors and course managers in case of any questions and difficulties. 6 core modules were selected and set up into the core body. They were equipped with guiding description, quiz or assignment, video lectures, evaluation questionnaire, folder of bibliography and resource. Quiz and assignment have been proven to be effective tools for inspiring engagement and should be set carefully, so as not to make learners feel self-contented or frustrated. In addition, it is the alternately sequenced modules that make learners obtain fresh experiences.

The last part, also called the end of the course, was endowed with the summary functions. The "Showcase Wall" or "Wrap-up Page" session functioned as a place for participants to share experiences and propose suggestions, share learning moments with photos/videos, and answer questions in the form of "building up a sequence". Other sessions like post-training questionnaire, course survey, and closing ceremony were also established in the last part. In fact, the ending part is crucial for the whole training course, for it provides an opportunity for recalling on what students have learnt, and gives them the sense of completion which drives motivation to the final sprint.

3.2 Social Collaboration Tools

Learners are easy to feel isolated and worn out after the long-time online studies, especially in the condition that no peers around physically. As a result, negative emotions may harm the online learning experiences or even end up in dropping out of courses, and therefore social activities are indispensable among trainees from time to time and they could highlight the engagement of the courses by properly planning and implementing with some outside collaborative tools. In this course, the Padlet [10] was introduced with the aim to build an extended "family bond" among all the participants including course managers and instructors, and encourage all the participants to enhance the "familiar sense" with each other.

An invisible "storyline" was designed to increase the interest for this community, and new tasks would be announced every few days, such as ice breaking, self-introduction, "coffee break", cultural exchange and festival introduction. Participants would get used to visit the pages to know about the updates by other learners during the whole training period without emotional tension. In that case, learners had to visit the online course modules as well, and formed imperceptible habits of learning.

Forum worked as another kind of asynchronous social learning tool in conducting online training activities, where students preferred posting and answering professional knowledge questions to casual chatting. That kind of module was advantageous to build discussion groups on some topics, and promote distant collaborations and mutual supports. Unfortunately, online chat room, as a typical synchronous tool, did not work efficiently enough for this practice, because it resulted in low attendance and high frequency of showing up and leaving, due to the wide distribution of time zones. Therefore, it is better to make a bundled solution of synchronous tools with other asynchronous modules in the case of online international education and training.

3.3 Learner Focus Services

Learner focus service is a kind of guarantee mechanism for increasing engagement and improving learning experiences, because students will feel cared about and reduce loneliness.

Learning experience online is sensitive to time delay. Unlike face-to-face training courses, most of the activities in international online courses are conducted in the environment of Internet, through which information may take considerable hops to the terminal equipment from the server, and the Internet conditions of the learners can be considerably variable. Sometimes, learners experience slow streaming and feedback on the technical issues. The enthusiasm for learning online could disappear after several fruitless attempts. For the problems above, it is necessary to carry out system monitoring and bug fixing, and an embedded folder for sharing files in the course is considered as a prompt backup solution, in case of streaming quality problems caused by network instability.

A user guide was offered to participants at the very beginning of the online course, explaining the technologies and skills needed during the whole training process, in order to avoid troubles caused by any technical problems. At the same time, there was an opportunity for learners to take a quick glance at the course framework. And the electronic schedule was also advertised, in order to make students know about the course arrangement without missing any required modules. In addition, annotation was suggested to be embedded into all the important modules, so that students could notice and follow the instructions. Besides, announcement in forum, information reminder and email were adopted, in order to make sure important messages were well received.

4 Course Evaluation, Deficiencies and Conclusion

According to the WMO-No. 1114 [1], course evaluation provides as a way of measuring the worth of providing a learning opportunity with the aim of improving the learning process by judging whether it met the objectives. And this course received positive results in both degrees of satisfaction and participation. On the other hand, several deficiencies emerged after analyzing qualitative feedbacks from participants, instructors and course managers.

4.1 Course Evaluation

In the case of online international course, learning outcome usually has a close coupling correlation with two measuring degree factors, which are satisfaction (emotional response) and participation (behavioral response). It means that if the course receives more approving comments and more activity involvements, it will possess more power of engagement.

This course evaluation was realized by setting up post-training questionnaires and collecting anonymous feedback data by participants with the system inbuilt survey and feedback modules before the end of the course. 118 post-training anonymous submissions were received and data was collected, and the degree of satisfaction was measured by the proportions of feedbacks in 5 levels, which are strongly disagree, disagree, neutral, agree, and strongly agree. As shown in Table 1, 91% of participants believed the training curriculum of the present training course was reasonable, 83% of participants thought the training method was effective for learning, 82% of participants approved that the teaching materials were suitable for learning, and 87% of participants agreed that the training management (such as the organizers for the training) was satisfactory.

Table 1. Result of online course feedback by degree of satisfaction.

Degree of satisfaction	Strongly disagree	Disagree	Neutral	Agree	Strongly agree
The training curriculum of the present training course was reasonable	0.00%	2.54%	5.93%	64.41%	27.12%
The training method was effective for learning	0.85%	4.24%	11.86%	52.54%	30.51%
The teaching materials were suitable for learning	2.54%	3.39%	11.86%	52.54%	29.66%
The training management (Such as the organizers for the training) was satisfactory	0.85%	0.85%	11.02%	58.47%	28.81%

Taking the trial course in 2018 as a control group, around 86% of learners felt the course practical and useful for work, whereas the figure from control group was 80%. Table 2 illustrates an obvious growth of involvement in countries and/or regions, registered participants and number of trainees who have 60% completion and above. Consequently, there is a steady increase in trainees and improvement in learning outcomes.

Table 2. Result of online course feedback by degree of participation.

Degree of satisfaction	Number of countries and/or regions	Registered participants	Number of trainees who have 60% completion and above
Study group	74	305	127
Control group	55	274	116

Learning experience describes the transaction that takes place between individual learners and the instructional environment. In addition to the components listed above, learning experience includes the way that the learner feels about, engages with, responds to, influences, and draws from the instructional situation [2]. In order to help understand the value in the online learning environment, another survey was also established and opened to all the participants. The survey offered 24 statements asking about preferred (ideal) experience without "right" or "wrong" answers. The opinions from participants were treated with a high degree of confidentiality, and would not affect their assessments. Relevance, reflective thinking, interactivity, tutor support, peer support and interpretation are designed as measurement scales for the course, and integer experience degrees from 0 to 5 represent "almost never", "seldom", "sometimes", "often", and "almost always" respectively.

Table 3 analyses 116 responses from trainees who were asked about the gaps between their ideals and senses of reality. In the chart, relevance is recognized to an "often" level, for the reasons like: it is important to and improved their practice, and more or less focus on interesting issues. Followed by it, reflective thinking, tutor support and interpretation are noted between "often" and "sometimes".

Table 3. Result of online learning survey by participants' experience.

Degree of experience	Relevance	Reflective thinking	Interactivity	Tutor support	Peer support	Interpretation
Average	4.06	3.53	3.02	3.75	2.87	3.58
Maximum	4.18	3.76	3.08	3.84	2.92	3.85
Minimum	3.94	3.26	2.94	3.68	2.74	3.23

Through further feedback analysis of data and comments, the learners of the course are often critical of their learning, own ideas and reading, sometimes they are critical on others. On the other hand, tutors often stimulate thinking, model discourse, and sometimes they encourage trainees or model self-reflection. In the aspect of interpretation, trainees are usually confident in understanding the tutors and other students, but they are not sure of being comprehended of what they would like to express. Interactivity and peer support are located at the level of "sometimes", which indicate that the collaboration and interaction are inadequate to well engage the learners, and course managers need to consider adding more collaborative and interactive ingredients to the modules. On the interactivity issue, students sometimes explain their ideas and ask for explanations, but they often receive few responses. Peer support is considered as a primary supplement on condition that instructors and facilitators are not available for a large group of students, and the survey shows the source of peer support sometimes are encourage, praise and value from other student, but empathizing is rarely adopted.

4.2 Deficiencies and Future Improvements

It seems difficult to have both advantages in training scale and training effect, which means trainees may not be able to complete their courses by the expected deadline.

Figure 3 illustrates the relation between trainees and their course completions, and 40% of them have completed at least 60% of the course which is almost close to the control group. And large training scale has also brought too much extra works to the course managers and instructors, who could not be able to concentrate enough on the implementation of instructional design. In addition, the teachers-student ratio is relatively low. Consequently, the next step is to thoroughly identify and analysis the number and competencies of trainees that is ideal for different levels of online international courses in the same subject.

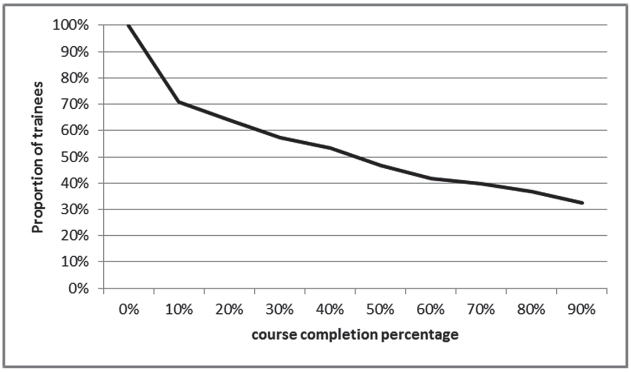


Fig. 3. The relation of trainees and their course completion.

In addition, some questions from the quiz and assignment modules have been reported as difficult by some of the trainees, and they more or less make learners reduce the course engagement, although the modules are announced only for self-practice, and would not account to the final score, or represent any of study status. In the future, it is essential to build a question bank objected at different levels for learner in order to increase more engagement.

4.3 Conclusion

The attempt to implement a series of online international training courses due to the threat of epidemic disease indicates that it is feasible and practical to turn physical classroom courses into online courses. However, this kind of conversion is not simply recording and sharing lectures from physical classroom. And some proper instructional designs that meet the needs and obey the cognitive rules of learners, will improve the learning experience and teaching effectiveness. The flexible integration of interactive modules into the whole teaching process has been proven to be effective in supporting the communication, collaboration and cooperation among large group of learners, through the practice of the International Distance Training Course on Short-term Climate Monitoring and Prediction in Disaster Prevention and Mitigation. And the implementation of information technology and innovation assists the teaching activities, and is conducive to improving learner's engagement and further the learning experiences online.

References

1. Guidelines for Trainers in Meteorological, Hydrological and Climate Services. Publications Board, World Meteorological Organization, Geneva, Switzerland, pp. 11–81 (2013). https://www.wmo.int/pages/prog/dra/documents/wmo_1114_en.pdf
2. Patrick, P.: Aesthetic principles for instructional design. Educ. Technol. Res. Dev. 57, 511–528 (2009). https://doi.org/10.1007/s11423-007-9060-7
3. Patrick, P.: Aesthetic decisions of instructors and instructional designers. In: Information Resources Management Association, Instructional Design: Concepts, Methodologies, Tools and Applications, pp. 1904–1920. IGI Global, Hershey (2011). https://doi.org/10.4018/978-1-60960-503-2.ch806
4. Patrick, P.: Plotting a learning experience. In: Luca, B., Todd, S. (eds.) Handbook of Visual Languages for Instructional Design, pp. 90–110. IGI Global, Hershey (2008). https://doi.org/10.4018/978-1-59904-729-4.ch006
5. Duncan, J.W., Steven, H.S.: Collective dynamics of 'small world' network. In: The Structure and Dynamics of Networks, pp. 301–303. Princeton University Press, Princeton (2011). https://doi.org/10.1515/9781400841356.301
6. Barabasi, A.-L., Albert, R.: Emergence of scaling in random networks. In: The Structure and Dynamics of Networks, pp. 349–352. Princeton University Press, Princeton (2011). https://doi.org/10.1515/9781400841356.349
7. Patrick, P., Tim, S., Joe, L., Wendy A.: The CALMet conference: international sharing of approaches to education and training in hydrology and meteorology. UCAR/COMET, Boulder, CO (2008). https://www.comet.ucar.edu/presentations/Abstracts_2008_2009/TheCALMetConference_PParrishAMSJan09.htm
8. Visualizing communities using 2D Pivot MDS. http://mrvar.fdv.uni-lj.si/pajek/community/DrawPivotMDS.htm. Accessed 28 Sept 2020
9. Moodle. https://moodle.org/. Accessed 28 Sept 2020
10. Padlet. https://padlet.com/. Accessed 28 Sept 2020

Improving Productive Classroom Talk Through Visual Learning Analytics Technology: A Case Study of an Award-Winning Mathematics Teacher

Chung Kwan Lo[1](✉) and Gaowei Chen[2]

[1] The Education University of Hong Kong, Hong Kong, China
chungkwanlo@eduhk.hk
[2] The University of Hong Kong, Hong Kong, China

Abstract. Visual learning analytics (VLA) technology is an emerging method to improve teaching and learning. However, few studies have examined the use of this technology in teacher professional development (PD). We thus conducted a case study to investigate the learning experience of an award-winning teacher in our year-long PD program for secondary school mathematics teachers in Shanghai. In our workshops on productive classroom talk, we used VLA technology to visualize the teacher's classroom instruction data from lesson videos, to support her teaching reflection. Data from lesson videos, teacher reflections, and an interview were collected and analyzed. The findings suggest that the teacher gradually increased her use of productive talk moves (e.g., press for reasoning) to elicit student responses, and used VLA to obtain evidence to support her lesson observations, teaching reflections, and refinements of her classroom instruction. Implications for the future practice of video-based VLA-supported PD activities are discussed.

Keywords: Visual learning analytics · Teacher-student interactions · Productive classroom talk · Professional development · Mathematics education

1 Introduction

Videos are increasingly used in teacher professional development (PD) [1, 2], because they can capture the richness and complexity of classroom activities for teachers to review. Although video viewing has clear benefits, the richness of video data can be overwhelming, and this richness can hinder teachers from noticing relevant events for reflection [3]. Therefore, we attempted to incorporate visual learning analytics (VLA) technology into our PD program for secondary school mathematics teachers to support their video viewing and reflection on their classroom instruction. VLA technologies represent educational data using various visualization strategies (e.g., bar plots and pie charts) to help teachers gain insight into the data [4]. However, few studies explore how such technologies can be used to (1) visualize data from a brick-and-mortar classroom setting and (2) facilitate teachers' improvements to their classroom instruction [2, 4].

L.-K. Lee et al. (Eds.): ICTE 2020, CCIS 1302, pp. 213–224, 2020.
https://doi.org/10.1007/978-981-33-4594-2_18

To remedy these gaps in previous knowledge, we present a case study conducted to investigate whether and how the use of VLA technology in our PD program changed a teacher's instructional approach. One approach to case selection in education research is to study unique cases [5], such as successful teachers [6]. This study thus focuses on the case of an award-winning teacher. We examine the changes in her classroom instruction and seek to understand her learning experience during our PD program. To confirm the effectiveness of her teaching, we further evaluate her students' learning gains using a pre-test and post-test. The findings of this case study may extend our understanding of how teacher educators can use VLA technology to improve their classroom instruction. This study thus lays substantial groundwork for the future practice of video-based VLA-supported PD activities. The following research questions (RQ1 to RQ3) are posed to guide this study.

1. How did our PD program influence the focal teacher's classroom instruction?
2. To what extent did the program help the teacher to increase her students' mathematics achievement?
3. How did the teacher use VLA technology to improve her classroom instruction?

2 Conceptual Framework of the PD Program

Our PD program leveraged three areas of the literature. First, we used an extended Vygotskian perspective on knowledge acquisition [7]. Second, the PD content drew on research concerning productive classroom talk [8]. Third, VLA technology was used to support our teacher participants' video viewing and reflection.

2.1 Extended Vygotskian Perspective on Knowledge Acquisition

Vygotsky [9] emphasizes the roles of teachers and capable others in knowledge acquisition. The Vygotskian perspective assumes that the teacher's job is to transmit knowledge and skills to learners through verbal instruction. However, Hatano [7] points out that this conventional perspective treats the learners as passive by nature. He therefore proposes extending the Vygotskian perspective by incorporating a co-constructivist view of teaching and learning. In line with Hatano's [7] idea, the inquiry mathematics approach proposed by Cobb et al. [10] aims to facilitate students' construction of mathematics knowledge. These researchers' teacher-guided class discussions encourage students to build on each other's explanations. Students are thus able to negotiate and co-construct ideas with their peers through dialog. As Hatano [7] concludes, teachers still play an important role in the knowledge acquisition process, but they should encourage their students to more actively construct knowledge.

2.2 Productive Classroom Talk

Previous research provides considerable evidence to support the extended Vygotskian perspective. In their synthesis of research on mathematics classroom talk, Kyriacou and Issitt [11] find that students have better learning outcomes when the teachers use questions to elicit correct answers, and then press the students for their reasoning and explanations. Resnick et al. [8] identify eight types of productive talk moves that teachers can use to probe students' thinking and engage them in productive classroom discussions. A recent PD program by Amador [1] applies some of these talk moves to develop the practice of pre-service mathematics teachers. Most importantly, these talk moves can improve student learning in mathematics and reasoning (Resnick et al. 2018). Therefore, our PD program introduced teachers to the productive talk moves proposed by Resnick et al. [8], as shown in Table 1. These moves can enrich teachers' knowledge of and skills in classroom instruction.

Table 1. Productive talk moves (summarized from Resnick et al. [8], p. 180).

Productive talk moves	Description	Examples
Say more	Ask a student for further elaboration	How did you arrive at that answer? Say more about that
Press for reasoning	Ask a student to explain his/her reasoning	Why do you think that?
Re-voice	Re-voice student expressions to elicit further responses	Let me see if I've got your thinking right. You're saying XXX?
Challenge	Challenge a student's idea	Can you think of any examples that would not work?
Restate	Ask students to restate someone else's reasoning	Can you repeat what he just said in your own words?
Agree/disagree	Ask students to state their standpoint	Do you agree or disagree?
Explain other	Ask students to explain someone else's reasoning	Why do you think he said that?
Add on	Prompt students for further participation	Would someone like to add on?

2.3 Video Viewing and VLA Technology

In recent years, teacher educators have shown increasing interest in providing video-based training [2]. With videos, teachers can review their lessons multiple times, which enables a deeper level of analysis of classroom instruction [12]. However, van den Bogert et al. [13] discover that some teachers tend to misinterpret classroom events and even to overlook them when watching the videos. According to Vieira et al. [4], the visualization of learning data or VLA can help to enhance teachers' selective attention to exchanges that are worthy of review. In the context of teacher PD, Chen [14] develop a classroom discourse analyzer that can provide teachers with VLA on their lesson videos. First, the teacher educators transcribe and code the dialog from the video lessons. The analyzer can then let reviewers visualize the dialog using (1) timelines of the teacher's and individual students' exchanges, and (2) speech bubbles of various sizes to represent the numbers of words used in each turn (see Fig. 3). Information on the classroom dialog, such as the frequency of each productive talk move used, is presented in graphs. Furthermore, teachers can click on the graphs to view the details of the corresponding video segments and transcripts. The visual representations, transcripts, and video footage are synchronized (see Fig. 2) to support the teachers' reflection when they watch the lesson videos.

3 Methods

As noted previously, this study aims to investigate how VLA technology can be used to support teacher reflection and to improve classroom instruction. A case study approach is most suitable to address our research questions, which require an extensive and in-depth description of the focal teacher's learning processes and outcomes [15].

3.1 Research Context and Case Selection

The case study was conducted during our year-long PD program for secondary school mathematics teachers in Shanghai. We recruited more than 50 teachers from 16 local secondary schools. Our case study focused on a first-year award-winning teacher. This teacher had majored in mathematics education and was serving as a mathematics teacher in a local secondary school (with around 2,600 students and 170 teaching staff). She taught a class of 24 Junior Secondary 2 students, aged from 12 to 13. In the middle and toward the end of our program, she received awards from the local government in recognition of her outstanding performance in teaching and lesson design. Looking at this unique case and learning from this successful teacher can provide insights into the future practice of video-based VLA-supported PD activities [5, 6].

3.2 Description of the PD Program

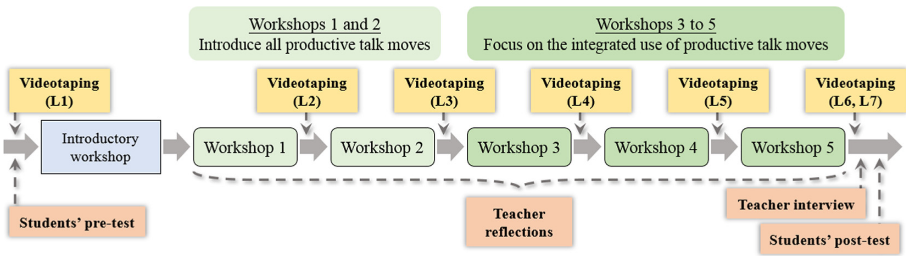


Fig. 1. Overview of the PD program and data collection. (Note: L1 means the first lesson of videotaping; L2 means the second; and so on.)

As shown in Fig. 1, our PD program offered one introductory workshop and five training workshops, conducted every one or two months throughout a year. Before attending the introductory workshop, our teacher participants videotaped their lessons. This workshop provided an overview of productive classroom discussions and the logistics of our PD program, on topics such as video submission requirements and workshop arrangements. After the workshop, the teachers submitted their lesson videos to us. Our research team then analyzed their videos and visualized the data using the VLA technology [14], as a resource to support the teachers' lesson reviews in the next workshop.

In each subsequent workshop, we first provided a short lecture on the PD content (Workshops 1 and 2: Introduction to productive talk moves; Workshops 3 to 5: Integrated use of productive talk moves). Our teacher participants then individually reviewed their own lesson videos and VLA. After that, they discussed their teaching performance in groups. Using the VLA technology, they could navigate various kinds of classroom instruction data, such as the teachers' production talk moves used in the lessons. In addition, they could review the transcripts of some specific video segments for in-depth reflection (see Fig. 2). Toward the end of each workshop, we engaged all of the teacher participants in a whole-class discussion, during which we shared a few good examples of different productive talk moves and highlighted some common problems to be dealt with in future practice. The teachers then completed written reflection forms to consolidate their learning, and to articulate their ideas on how to improve their future practice. After each workshop, the teachers videotaped their lessons and submitted their videos to us again.

3.3 Data Collection and Analysis

Using multiple data sources allowed us to triangulate the data, which could strengthen the validity of our case study [15]. Data were collected from four major sources, namely lesson videos, students' pre-test and post-test, teacher reflections, and a teacher interview.

Lesson Videos. To examine our award-winning focal teacher's changes in classroom instruction (RQ1), her 40 min lessons were videotaped and analyzed every one to two months. We focused on the lessons of teacher-led whole-class discussions on new subject contents, during which the teacher and students interacted (i.e., teacher–student and student–student) in front of the whole class. Figure 1 shows the schedule of videotaping (L1 to L7). Our research team transcribed her lesson videos (a total of about 280 min and 42,000 words in Chinese characters from the teacher and her students), and then coded the classroom discourse into the eight types of productive talk moves (Table 1). The transcripts were double coded by two of our research team members. Any disagreements in coding were resolved through discussion to reach a consensus. In this study, we present our analysis of (1) the numbers of the productive talk moves that the teacher used, and (2) the percentages of the teacher's and her students' word contributions per lesson. The influence of our PD program could thus be revealed by tracking these statistics across the videos recorded at different times.

Student Tests. To evaluate the students' learning gains after a year of mathematics learning (RQ2), we conducted a 40 min pre-test and post-test. To ensure their validity and reliability, these tests were designed on the basis of the Trends in International Mathematics and Science Study (TIMSS), and the two tests were similar in terms of their scope and difficulty. The total score in each test was 47, and they each covered the three cognitive domains of knowing (9 marks), applying (15 marks), and reasoning (23 marks). Our preliminary exploration of the test data discovered a significant deviation from normality in the post-test data ($p < .001$). Therefore, the Wilcoxon signed-rank test—a non-parametric test to compare two dependent samples [16]—was run to analyze the test data. The effect size (r) was calculated using the formula provided by Field [16], with the benchmarks of .3 and .5 representing medium and large effect sizes, respectively.

Teacher Reflection and Interview. To gain an in-depth understanding of the focal teacher's perceptions of our video-based VLA-supported PD activities and of how she used the analytics to improve her classroom instruction (RQ3), we collected teacher reflection and interview data. At the end of each workshop, the teacher completed a written reflection form, which recorded her immediate thoughts on what she had learned and her plan to improve future practice. Guiding questions included "In this workshop, what have I learned about how to facilitate students' thinking through classroom talk?" and "How can I further improve the classroom instruction of my mathematics lessons?" In addition, an in-depth interview was conducted at the end of our PD program. Some sample questions of our interview protocol were "How do you perceive the use of VLA technology in our PD program?" and "How do you improve your classroom instruction based on your analytics of lesson videos?"

The teacher reflection and interview data were analyzed using a series of qualitative data analysis procedures. The data were first transcribed in Chinese. We then identified patterns in the data [15]. Some of the data were translated into English for reporting purposes. To establish coding reliability, the data were double-coded by the two authors. Any disagreements were resolved through discussion. To enhance validity, we asked the teacher to read and confirm the accuracy of the reflection and interview transcriptions (i.e., member checking), and we reported our findings using direct quotations.

4 Findings

4.1 Teacher's Instructional Changes (RQ1)

We analyzed the changes in the teacher's classroom instruction over time. The two major aspects we focused on were (1) the number of the teacher's productive talk moves used and (2) the percentages of the teacher's and her students' word contributions in the lessons. First, Table 2 shows that the number of productive talk moves that the teacher used in her lessons gradually increased from fewer than 10 (L1) to over 40 (L4 to L7). Furthermore, following the completion of Workshop 2 (L4 onward), the variety of her productive talk moves increased. For example, in L4 she started to use the "agree/disagree" and "explain other" talk moves.

Of the 249 identified productive talk moves from L1 to L7, the teacher most frequently used "say more" (47.8%), "press for reasoning" (22.9%), and "add on" (17.7%). Figure 2 shows a discourse excerpt example of "press for reasoning," as displayed in our classroom discourse analyzer. However, the teacher rarely used the five other types of productive talk moves, such as "explain other" (2.0%) or "restate" (0.8%).

Table 2. Number of productive talk moves used per lesson over time.

	L1	L2	L3	L4	L5	L6	L7	Overall
Say more	5 (55.6%)	12 (52.2%)	5 (17.9%)	24 (43.6%)	26 (54.2%)	15 (36.6%)	32 (71.1%)	119 (47.8%)
Press for reasoning	4 (44.4%)	3 (13.0%)	16 (57.1%)	12 (21.8%)	4 (8.3%)	12 (29.3%)	6 (13.3%)	57 (22.9%)
Re-voice			1 (3.6%)	1 (1.8%)	3 (6.3%)	1 (2.4%)	1 (2.2%)	7 (2.8%)
Challenge			1 (3.6%)	4 (7.3%)	1 (2.1%)	2 (4.9%)		8 (3.2%)
Restate					2 (4.2%)			2 (0.8%)
Agree/disagree				4 (7.3%)	1 (2.1%)	1 (2.4%)	1 (2.2%)	7 (2.8%)
Explain other				4 (7.3%)			1 (2.2%)	5 (2.0%)
Add on		8 (34.8%)	5 (17.9%)	6 (10.9%)	11 (22.9%)	10 (24.4%)	4 (8.9%)	44 (17.7%)
Total	9	23	28	55	48	41	45	249

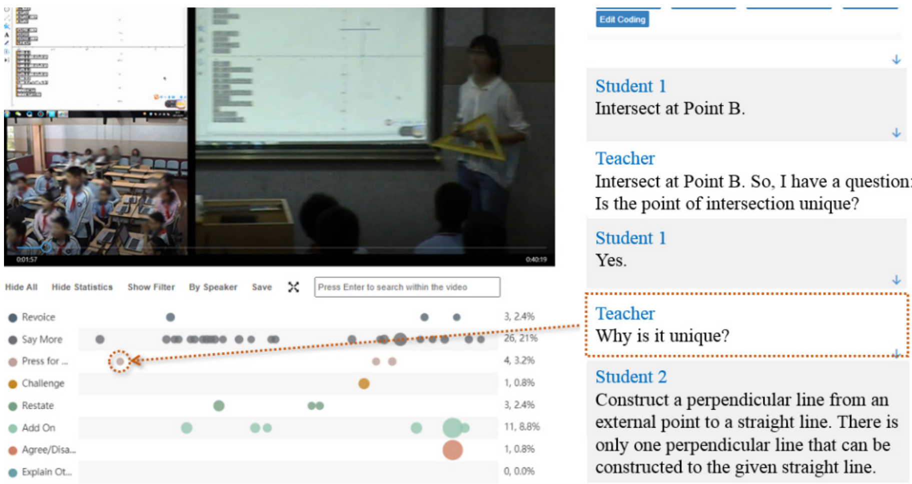


Fig. 2. A discourse excerpt example of "press for reasoning." (Note: The excerpt is translated into English for reporting purposes.)

Second, the total number of word contributions (i.e., teacher + students) per lesson ranged from 4,831 to 7,021 ($M = 5{,}973$; $SD = 953$). Table 3 shows that the percentage of the teacher's word contributions per lesson decreased from over 85% (L1 and L2) to about 70% (L6 and L7). Correspondingly, the percentages of her students' word contributions per lesson increased from about 15% to 30% over time.

Table 3. The number of the teacher's and the students' word contributions per lesson over time.

	L1	L2	L3	L4	L5	L6	L7
Teacher	4,894 (89.8%)	6,082 (86.6%)	5,839 84.0%)	4,417 (83.0%)	5,727 (82.5%)	3,227 (66.8%)	3,726 (70.3%)
Students	556 (10.2%)	939 (13.4%)	1,109 (16.0%)	904 (17.0%)	1,215 (17.5%)	1,604 (33.2%)	1,573 (29.7%)

4.2 Students' Learning Gains in Mathematics Lessons (RQ2)

The students' learning gains were evaluated using a pre-test and post-test. Overall, the results of the Wilcoxon signed-rank test indicated that the students (n = 24) scored significantly higher in their post-test (*Mdn* = 39.00) than their pre-test (*Mdn* = 30.50), $z = 3.69$, $p < .001$, with a large effect size of $r = .53$. Table 4 further shows that their learning gains in all three domains were significant, with a large effect for knowing ($r = .61$) and median effects for applying ($r = .46$) and reasoning ($r = .43$).

Table 4. Results of students' pre-test and post-test.

		Pre-test		Post-test			
Domain	Max	*M* (*SD*)	*Mdn*	*M* (*SD*)	*Mdn*	z	r
Knowing	9	5.37 (1.61)	5.00	8.83 (.82)	9.00	4.22***	.61
Applying	15	9.50 (3.32)	11.00	12.21 (2.43)	13.00	3.20**	.46
Reasoning	23	15.13 (1.65)	15.00	17.08 (3.43)	18.00	2.98**	.43
Total	47	30.00 (4.92)	30.50	38.13 (6.23)	39.00	3.69***	.53

**$p < .01$; ***$p < .001$

4.3 The Teacher's Ways of Using VLA (RQ3)

The teacher reflection and interview data suggest that the teacher used VLA of her lesson videos to support video viewing. More specifically, the VLA technology provided evidence that helped her to *observe* her lessons. She then *reflected* on her teaching performance and *refined* her approaches for engaging her students in productive classroom discussions. We report the following two episodes to illustrate her pattern of VLA-supported lesson review.

Using More Productive Talk Moves. As shown in Table 2, the teacher used more productive talk moves to elicit student responses after training. She asserted that "The discourse analyzer platform can make me clearly understand the classroom instruction in my mathematics lessons" (i.e., observation). After reviewing the classroom dialog in her lessons, she realized that her requests for choral responses or short answers were not effective, because the students tended simply to confirm the teacher's statement without deep thinking (i.e., reflection). As she shared,

When I worked out a mathematics proof, read it out, and then asked students: "Is that right?" I found that such a question was not effective, because the students would basically say "yes." But actually, they would not think too much.

Therefore, she attempted to use more productive talk moves such as "say more" and "add on" to engage the students in productive classroom discussions (i.e., refinement). In her words,

Referring to my analytics, I find that encouraging students to say more or asking other students to add on to one's idea are better [moves] … When students voice their ideas, no matter right or wrong, other students are able to judge. At the same time, they can consolidate their thinking and even identify their own mistakes.

Fig. 3. A comparison of the VLA of (a) L1 and (b) L7 by speaker.

Speaking Less and Allowing Time for Student Thinking. As shown in Table 3, the percentages of the students' word contributions per lesson increased over time. The teacher shared that the VLA had a great impact on her understanding of teacher–student interactions during the lessons (i.e., observation). Figure 3(a) shows a series of large speech bubbles in the teacher timeline, alongside a few scattered small speech bubbles in the students' timelines in L1. When reviewing this lesson video, the teacher realized that her lesson was teacher-dominated. Most importantly, she found that most of her talk turns were not meaningful (i.e., reflection). In her words,

I think your analyzer had a great impact on my understanding of my own teaching. At a glance, I realized that I spoke so much nonsense at the start of your program—that is, the first lesson. A series of large speech bubbles! I spoke a lot in each turn.

To improve, she attempted to speak less and allow more time for her students to think and formulate their responses (i.e., refinement). As she mentioned, After viewing this [the VLA of her lesson videos], I tried to speak less and asked more questions … When my students did not respond, I would speak more slowly and allow more time for them to think. Therefore, my word contributions became fewer and the students explained more.

In L7 (Fig. 3(b)), compared with L1, the number of students' speech bubbles (i.e., the number of students' talk turns) and their sizes (i.e., the numbers of word contributions) generally increased. In other words, the teacher was better able to elicit detailed student responses after her reflection and refinement of her teaching practice.

5 Discussion

Using lesson videos in PD activities does not necessarily facilitate teacher reflection and instructional improvement [3, 13]. This problem raises the question of how to enhance teachers' efficacy in video viewing and in noticing relevant events that are worthy of a review. We used VLA technology to visualize the classroom instruction data from lesson videos. We studied how a first-year award-winning teacher made use of VLA in video-based PD activities. Our findings suggest that there was an improvement in her teaching practice in terms of the use of productive talk moves and the students' word contributions in the lessons. In addition, the results of the student tests indicated a significant learning gain after one year of mathematics learning.

In this case study, we identified a pattern in the teacher's video-based VLA-supported lesson reviews. She first observed her lesson with a focus on the use of productive talk moves and on her students' performance in classroom discussions. Next, she reflected on her classroom instruction and refined her approaches for engaging the students in productive classroom discussions. As Chow et al. [17] explain this process, she thought and worked like an action researcher. If teacher educators can empower teachers to become researchers of their own teaching practice, then teacher professionalism can be enhanced [17]. Based on the lessons that we learned from this case study and the methodological work of Kemmis and McTaggart [18], we suggest the following three stages for guiding teachers in the systematic review of video lessons, as they practice VLA-supported PD activities.

First, the teachers have to observe their own teaching practice [18]. Teacher educators can guide them in observing their lessons according to two criteria: (1) the overall use of productive talk moves, and (2) the percentages of the teacher's and the students' word contributions per lesson. By referring to the timeline and speech bubbles of the participants, the teachers can gain an overall idea of the teacher–student interactions during the lesson. In addition, teacher educators can ask the teachers to compare the VLA of their current lessons with those of their previous lessons. In this way, the teachers can monitor their instructional changes and improvements. Some possible guiding questions are "What kind of productive talk moves have I used most frequently?", "What kind of productive talk moves have I used the least?", and "Compared with my previous lessons, is there an increase in the students' word contributions?"

Second, teachers have to reflect on their teaching practice [18]. According to Resnick et al. [8], each type of productive talk move has its own value. Furthermore, classroom discussions are more engaging if the teachers can use various types of productive talk moves to stimulate their students' thinking and elicit their responses. Therefore, teacher educators can guide the teachers to reflect on their distributions of productive talk moves in their lessons. Furthermore, teacher educators can ask the teachers to review each individual talk move by viewing their video segments and transcripts. With reference to student responses, the teachers can evaluate whether their talk moves are appropriately formulated for their own classes. Some possible guiding questions are "Why did I seldom use some types of productive talk moves?", "Do I understand how to use these productive talk moves?", and "Are my questions too difficult for my students?"

Finally, teachers have to refine their teaching practice [18]. After the observation and reflection stages, teacher educators can ask the teachers to identify aspects that they need to improve. If we take the teacher in this case study as an example, we can see that although she was able to allow more time for her students to voice their ideas after training, Table 2 shows that she seldom used certain types of productive talk moves, such as "challenges." However, this talk move is particularly important in mathematics learning. By providing appropriate challenges, teachers can extend their students' existing thinking and provide alternative pathways to understanding mathematics concepts [19]. On this point, regarding the limited use of some talk moves, an improvement was required in the teacher's practice. Some possible guiding questions are "Which aspects of teaching practice do I need to improve?", "How can I shorten my instruction to free up class time for students' contributions?", and "How can I use other types of productive talk moves to stimulate student thinking?"

6 Conclusion

This study has important implications for teacher educators. We demonstrate how VLA technology can be used to visualize lesson video data. We also provide rich insights into how an award-winning mathematics teacher made use of this VLA technology to improve her classroom instruction. VLA technology can make teachers' video viewing more purposeful. For example, the teacher featured in this study focused on the productive talk moves used in her lessons, and refined her approaches to eliciting student responses. Further efforts should incorporate speech recognition technology which can lessen the amount of our manual work (e.g., transcription).

To help empower teachers as researchers, our study adds to the literature by establishing three stages of helping teachers to use VLA of their lesson videos as evidence in their lesson observations, reflections, and refinements of their teaching practice. Although this case study lays substantial groundwork for the future practice of video-based VLA-supported PD activities, further large-scale studies are required to test and improve the use of this PD framework. Besides quantitative evaluation (e.g., number of word contributions), qualitative assessment of student responses can be future work for researchers. In addition, investigating the learning experiences of teachers with different backgrounds (e.g., experienced mathematics teachers or teachers from other subject disciplines) has the potential to enrich the proposed guidelines.

Acknowledgements. This research was supported by Hong Kong RGC (grant No. 27606915 and 17608318) and by the PPR Funding Scheme from the Policy Innovation and Co-ordination Office of Hong Kong SAR (project No. 2017.A8.073.18C).

References

1. Amador, J.M.: Video simulations to develop preservice mathematics teachers' discourse practices. Technol. Pedagog. Educ. **27**, 1–14 (2018)
2. Major, L., Watson, S.: Using video to support in-service teacher professional development: the state of the field, limitations and possibilities. Technol. Pedagog. Educ. **27**, 49–68 (2018)
3. Gaudin, C., Chaliès, S.: Video viewing in teacher education and professional development: a literature review. Educ. Res. Rev. **16**, 41–67 (2015)
4. Vieira, C., Parsons, P., Byrd, V.: Visual learning analytics of educational data: a systematic literature review and research agenda. Comput. Educ. **122**, 119–135 (2018)
5. Yin, R.K.: Case study methods. In: Green, J.L., Camilli, G., Elmore, P.B. (eds.) Handbook of Complementary Methods in Education Research, pp. 111–122. American Educational Research Association, Washington (2006)
6. Kumi-Yeboah, A., James, W.: Transformational teaching experience of a novice teacher: a narrative of an award-winning teacher. Adult Learn. **23**, 170–177 (2012)
7. Hatano, G.: Time to merge Vygotskian and constructivist conceptions of knowledge acquisition. In: Forman, E.A., Minick, N., Stone, C.A. (eds.) Contexts for learning: Sociocultural dynamics in children's development, pp. 153–166. Oxford University Press, New York (1993)
8. Resnick, L.B., Michaels, S., O'Connor, M.C.: How (well-structured) talk builds the mind. In: Preiss, D.D., Sternberg, R.J. (eds.) Innovations in Educational Psychology: Perspectives on Learning, Teaching, and Human Development, pp. 163–194. Springer, New York (2010)
9. Vygotsky, L.S.: Mind in Society: The Development of Higher Psychological Processes. Harvard University Press, Cambridge (1978)
10. Cobb, P., Wood, T., Yackel, E.: Discourse, mathematical thinking, and classroom practice. In: Forman, E.A., Minick, N., Stone, C.A. (eds.) Contexts for learning: sociocultural dynamics in children's development, pp. 91–119. Oxford University Press, New York (1993)
11. Kyriacou, C., Issitt, J.: What characterizes effective teacher–pupil dialogue to promote conceptual understanding in mathematics lessons in England in key stages 2 and 3? A systematic review. University of London, London (2008)
12. Khong, T.D.H., Saito, E., Gillies, R.M.: Key issues in productive classroom talk and interventions. Educ. Rev. **71**, 334–349 (2019)
13. van den Bogert, N., van Bruggen, J., Kostons, D., Jochems, W.: First steps into understanding teachers' visual perception of classroom events. Teach. Teach. Educ. **37**, 208–216 (2014)
14. Chen, G.: A visual learning analytics (VLA) approach to video-based teacher professional development: Impact on teachers' beliefs, self-efficacy, and classroom talk practice. Comput. Educ. **144**, 103670 (2020)
15. Yin, R.K.: Case Study Research and Applications: Design and Methods, 6th edn. SAGE, Thousand Oaks (2018)
16. Field, A.: Discovering Statistics Using SPSS, 3rd edn. SAGE, London (2009)
17. Chow, K.C.K., Chu, S.K.W., Tavares, N., Lee, C.W.Y.: Teachers as researchers: a discovery of their emerging role and impact through a school-university collaborative research. Brock Educ. J. **24**, 20–39 (2015)
18. Kemmis, S., McTaggart, R.: Participatory action research: communicative action and the public sphere. In: Denzin, N.K., Lincoln, Y.S. (eds.) The SAGE Handbook of Qualitative Research, 3rd edn., pp. 559–603. SAGE, Thousand Oaks (2005)
19. Anthony, G., Walshaw, M.: Effective pedagogy in mathematics. In: Brophy, J. (ed.) Educational Practices Series – 19. International Academy of Education, Switzerland (2009)

Open and Collaborative Learning

Supporting Social Regulation of Learning with Multi-dimensional Scaffolding in Computer-Supported Collaborative Writing Activities

Kailiang Chen[1], Haogang Bao[1], Yanyan Li[1](✉), You Su[1,2], and Yu Peng[1]

[1] School of Educational Technology, Beijing Normal University, Beijing, China
liyy@bnu.edu.cn

[2] School of Humanities, Beijing University of Posts and Telecommunications, No.19, Xinjiekouwai Street, Haidian District, Beijing, China

Abstract. Social regulation of learning has been recognized importantly in computer-supported collaborative learning (CSCL) context. The social regulation process is complex and requires technological scaffold. CSCL technologies such as script and group awareness tools have been utilized to support social regulation of learning. However, such tools are limited in supporting the planning and evaluating stage of regulation and failed to support the whole process of social regulation of learning. This research aims to design and develop a set of multi-dimensional scaffoldings to support planning, monitoring, regulating and evaluating stages of social regulation of learning based on learning analytics technologies. Moreover, the present work also investigates the validity of the proposed scaffoldings and students' acceptance of these tools.

Keywords: Computer-supported collaborative learning · Learning analysis · Multi-dimensional scaffolding

1 Introduction

Social regulation of learning has been recognized as one of the key factors that affect the learning performance in CSCL context (Järvelä and Hadwin, 2013; Su, et al. 2018). Social regulation of learning contains the regulation between group members (co-regulation of learning, CoRL) and the whole group (socially shared regulation of learning, SSRL) in cognition, emotion, motivation, and behavior (Zimmerman and Schunk 2011). As the social regulation process is complex and requires students' regulatory skills and strategies, studies have found that many learners experience challenges in applying successful regulation of learning in collaborative settings (Miller and Hadwin 2015). Scaffolding or supporting tools is required for facilitating the regulation of collaborative learning (Järvelä et al. 2016; Miller and Hadwin 2015). Several scripting and awareness tools have been successfully used in supporting the planning, evaluation, emotion awareness and engagement awareness of social regulation of learning. However, most

L.-K. Lee et al. (Eds.): ICTE 2020, CCIS 1302, pp. 227–238, 2020.
https://doi.org/10.1007/978-981-33-4594-2_19

of these tools support only planning stage and evaluation stage in social regulation of learning with predefined script. They failed to collect learning process data to support the whole process of social regulation of learning. Research advocates using learning analytics technology to build multi-dimensional support to scaffold social regulation of learning (Malmberg et al. 2017).

For designing effective learning scaffold for learners to develop their regulatory skills and knowledge in collaboration, the present study attempts to design and develop a set of multi-dimensional scaffolding, such as text description, promote questions, visual awareness information and peer feedback to, basing on the social regulation theories (Järvelä and Hadwin 2013; Su et al. 2018) and existing tools (Järvelä et al. 2016; Miller and Hadwin 2015). This research further investigates the usage of the proposed scaffolding tools and the students' acceptance of them.

2 Supporting Social Regulation of Learning in CSCL

In collaborative learning, the social regulation of learning emphasizes that learners can define collaborative learning goals, make learning plans, evaluate learning outcomes, and timely regulate learning in the learning process. social regulation of learning is made up of two types: co-regulation of learning and socially shared regulation of learning (Hadwin and Oshige 2011). They can influence the process of collaborative group discussion and influence the learner's cognitive construction process, which helps learners to complete tasks better.

In the current research, scaffold supported by social regulation of learning mainly includes three types: group awareness tools, collaborative script tools, and evaluation reflection tools. Among them, the group awareness tool refers to the information that the learner perceives the same group members, and provides the students with the knowledge, understanding or opinions of the collaborative group in the process of collaboration. It is divided into cognitive awareness tools and social awareness tools (Janssen and Bodemer 2013). Collaborative scripts are defined in the collaborative activities to define the task description of the group, the composition of the team, and the rough steps or specific processes of the collaboration, which are divided into macro scripts and micro-scripts. The evaluation reflection tool is to analyze the information related to the collaboration process or collaboration results, which is helpful for collaboration groups to evaluate themselves, peers and groups together. This study combs the research status of scaffolds supported by social regulation of learning, as shown in Table 1.

By comparing the typical scaffolds in Table 1, we can see that the current scaffolds have the following characteristics and disadvantages: first, the current scaffolds mainly support social regulation of learning from perception, script, and evaluation, but most scaffolds only focus on one or two aspects, and the completeness and systematizes of the scaffolds are insufficient; Second, the scaffolds are mainly aimed at supporting the planning and evaluation stages of collaborative learning, and lack of support for social regulation of learning in the process of collaborative learning. Third, the main data source of current social regulation's scaffold is learner's self-report or questionnaire, which can only guide and intervene social regulation of learning in stages and cannot support social regulation of learning dynamically and in real-time by using multi-source and multi-modal data such as homework, dialogue, logs, etc. Fourth, the current scaffolds mostly communicate in the form of forums and so on, which lack real-time interactivity and cannot provide a good learning environment for real-time collaborative editing and discussion.

Table 1. Scaffolding tools with CSCL technologies.

Tool name	Group awareness	Collaboration script	Evaluation reflection	Way of communication
REG (Laru et al. 2015)	Social awareness	Macro- script	√	–
Gstudy (Hadwin et al. 2010)	Cognitive awareness & Social awareness	Macro- script	√	gChat tool
collaboration script (Kim and Lim 2017)	Cognitive awareness	–	√	–
Nearpod Interactive Classroom Tool (Lai et al. 2018)	Cognitive awareness & Social awareness	–	√	Forum
Radar (Järvelä et al. 2015)	Cognitive awareness	–	√	–
Shared Planning Tool (Järvelä et al. 2016)	Cognitive awareness & Social awareness	–	–	–

In order to solve the above problems, the research combines learning analysis technology, designs and develops a set of comprehensive and systematic dynamic social regulation scaffolds to guide learners' cooperation and interaction according to the characteristics of group awareness, script and evaluation reflection, supports cognition, society and behavior, provides monitoring and feedback, and improves the quality of team collaborative.

3 Development of the Multi-dimensional Scaffold for Collaborative Writing

3.1 Learning Context and Environment

The scaffolding tools are designed for the synchronized English collaborating writing context. In this context, group students discuss and doing the writing task in the online synchronized writing environment face to face. Basing on the four-stage model of social regulation of learning (planning, monitoring, regulating and evaluating, Su et al. 2018), the process of collaborative learning activity mainly includes the following steps. In the first step, the learner completes the study plan of the individual and the collaborative group by combining the task understanding list given by the teacher in the task planning module. In the second step, the learner will enter the group interactive sharing module to perform the collaborative task with the group members. At the same time, learners can observe the visualization results of the group awareness module, so that learning strategy adjustments can be made in a timely manner. In the third step, after completing the collaborative task, the learner needs to enter the reflection evaluation module to score and comment on the learning outcomes of other collaborative groups. Finally, learners can improve the work of their group through feedback from other groups.

The integration of scaffolds follows the design principles of Järvelä et al. (2015), which are to scaffold social regulation, a suitable collaborating environment that contains the awareness, supporting externalization, and prompting function are needed. This environment is shown in Fig. 1.

Fig. 1. Learning environment.

3.2 Design of the Multi-dimensional Social Regulation Scaffold

Following this umbrella macro-scripted framework, we design multimodal micro-scripted level scaffolding strategies to support students' social regulation in different stages. Empowered by learning analytics technology (Zheng et al. 2018), these scaffolding tools can not only provide text guidance and promotes but also visual awareness information in cognitive, metacognitive and social aspects. Table 2 illustrates the multi-dimensional scaffolding tools that are integrated into the learning environment.

In social regulation, the purpose of scaffold planning stages is (1) building individual's task understanding, (2) promoting group's shared task perception and (3) engaging more planning discussion (Hadwin et al. 2018). Basing on previous work (Miller and Hadwin 2015; Splichal et al. 2018), we design text-based micro-script promoting questions and task descriptions to support students thoroughly analyze the task requirement and planning collaboration strategies, the planning tools are in Fig. 2.

Table 2. The multi-dimensional scaffolds for different social regulation.

Stages	Scaffolding	Functions	Tools
Planning	Task description Individual planning guidance Group planning guidance	Guiding students to understand the task, make the learning object, individual learning plan and share the task and assessment perception in the group,	Task description Promoting questions
Monitoring/Regulating	Cognitive awareness Metacognitive awareness Social awareness	Supporting group awareness of cognitive, metacognitive and social, so that students can monitoring the learning progress	Multiple Visualable interaction charts
Evaluating	Individual assessment Peer assessment	Providing evaluated questions and peer feedback to the student to refine their work before finish	Visual charts Peer feedback

Fig. 2. Individual planning and group planning.

In monitoring and regulating phase, the key of the scaffold is to support collaborators aware of individual or collective actions, task progress and social engagement, so that students can operate the information and adapt strategies to optimize the collaboration process (Miller and Hadwin 2015). Follow the advice of Järvelä and Hadwin (2013), we design "keyword cloud" as cognitive awareness to implicate the topic and words using of collaborative writing, "time reminder" and "word reminder" as metacognitive awareness for comparing the time and task completion, "group participation" as social awareness tool to frequency and intensity of students' communication. The awareness chart is shown in Fig. 3.

Fig. 3. Group awareness information.

The evaluation phase is helpful for team decision-making and adaptation of collaborative processes, progress, and products, thereby intentionally optimizing learning where needed (Zimmerman and Schunk 2011). It consists individual evaluation and peer assessment. Research has proved that combining peer assessment with group awareness tools is an effective strategy for reducing social loafing and ensuring fair assessments (Lin et al. 2019). As shown in the upper left corner of Fig. 3, in this work, we designed the chart and text awareness information for the individual and peer evaluation.

3.3 The System Architecture of Scaffolding Tools

In this study, Python, JavaScript, Ajax, and other technologies were used to develop the social regulation tool to support learners. The regulation scaffolds can not only collect log information such as the number of login and learning time of the students, but also analyze the text information, peer interaction information, posts and content, and evaluation in the collaborative discussion. Through statistical analysis, semantic analysis, keyword analysis, and other learning analytic methods, we analyze learners' social, behavioral and cognitive dimensions in the process of collaboration in a multi-dimensional way. The overall structure of the social regulation learning scaffolds can be logically divided into four layers. Each layer is connected by a data bus. The specific structure is shown in Fig. 4:

The data layer is the bottom layer of scaffolds, whose main function is to collect and store learners' learning process data. In order to better analyze the learning process.

The main function of the feature analysis layer is to process the underlying data, including three main steps of data pre-processing, Eigenvalue extraction and dimension aggregation. By analyzing the data of posting volume, interaction frequency, knowledge point mention frequency, social relationship density, and behavior type statistics, we can extract various social network indicators of knowledge point coverage to represent the learning process of individual learners and group learners.

The visual representation layer is to transfer various data reports from the previous stage into the visual presentation module for visualization processing.

Application presentation layer to provide learner-oriented analysis and monitoring services as application goals. Based on the analysis results of the visual representation layer, the application presentation layer provides different analysis results for the learner to help the user obtain the available information.

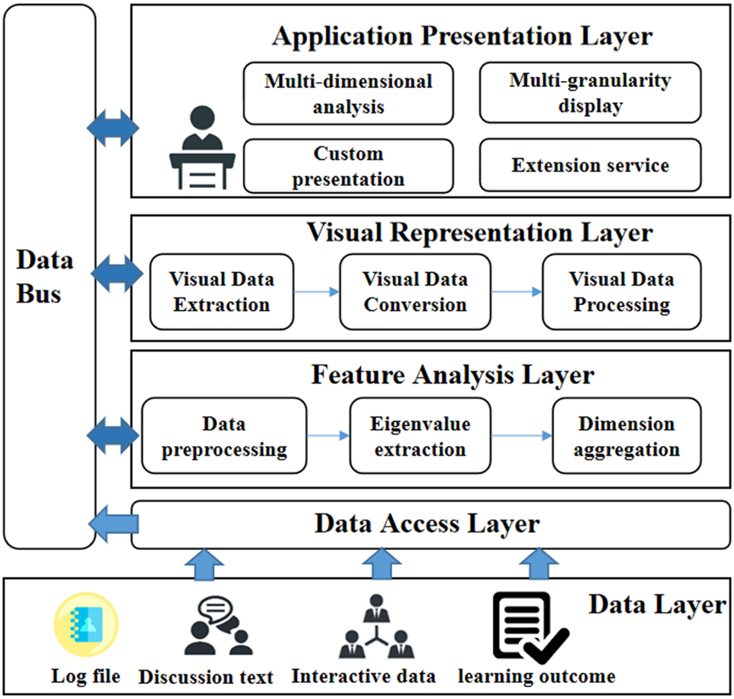


Fig. 4. The system architecture.

4 Method

4.1 Research Context and Participants

The activities of this study were carried out in the regulation learning scaffolds developed. Participants were 24 students from the second year of an undergraduate degree.

Collaborative learning activities are conducted in the "English Intensive" compulsory course offered by a university in China. Learners need to participate in a face-to-face traditional course of 100 minutes per week, according to the course schedule. In this course, students are freely divided into 7 groups of 3–4 people each. In the first 40 min of the course, the instructor will teach new knowledge and post collaborative learning tasks, so learners will spend 60 min on scaffolding tools to complete collaborative learning tasks.

4.2 Research Tool

In order to analyze the group's social regulation behavior in the process of online collaborative activities, this study encodes social regulation from two aspects: social regulation type and focus. This study adapted the coding table of Su et al. (2018) to form the coding table of social regulation behavior in this study. Among them, the form of adjustment is mainly peer adjustment and social collective adjustment, and the focus of adjustment mainly includes "Content monitoring", "task understanding", "process monitoring", "emotion" and "organization". Details of the coding description can be found in the original paper (Su et al. 2018).

4.3 Data Collection and Analysis

In this study, the social regulation behavior of learning groups is coded. Two researchers partitioned the learner's discussion records into effective coding units, and the consistency of the segmentation was referred to Chang et al.'s 2017) calculation method. The two researchers trained before the segmentation, and finally, the consistency of the two researchers was 0.87, greater than 0.8 to meet the requirements. After segmenting the discussion record, the two coders independently coded the learners' social adjustment behavior. The coding consistency coefficient Cohen's $k = 0.82$ in terms of "regulation form". In terms of "focus of regulation", the coding consistency coefficient Cohen's $k = 0.78$. The data show that the internal consistency of the two coders is acceptable.

5 Result and Discussion

5.1 Regulation Characteristics of Collaborative Groups Supported by Scaffolding Tools

In order to explore the role of the tool in this study, we analyzed the data of the learning group during the collaborative activities. We chose three of them as the members of the three groups participated in all the collaborative activities. The three groups are "Group 2", "Group 3" and "Group 5". According to our segmentation of the discussion record, we have counted 77 discourse fragments in the "Group 2", 31 discourse fragments in the "Group 3" and 78 discourse fragments in the "fifth group". As shown in Table 3.

In terms of the "Regulation type", we found that the more types of regulation in these three groups were socially shared regulation, and the co-regulation was less. It can be seen that in the collaborative learning activities, most learners tend to discuss the

Table 3. Characteristics of social regulation behavior.

Social regulation behavior	Frequency (Group 2)	Frequency (Group 3)	Frequency (Group 5)
Regulation type			
Co-regulation	36	14	33
Socially shared regulation	41	17	45
Regulation focus			
Content monitoring	43	13	27
Task understanding	9	8	14
Process monitoring	23	5	34
Organization	2	4	2
Emotion	0	1	1

problems encountered in the collaborative learning process with the members, reach a consensus, and finally formulate a solution to the problem. There are also some learners who will help other members of the group to adjust their learning, such as reminding a learner to modify the wrong content.

The results of this study show that socially shared regulation appears more frequently in "regulation type". This is the same result as Su et al. (2018) in the collaborative learning process, and the most common type of regulation is a socially shared regulation. This result is consistent with the characteristics of collaborative learning, that is, learners participate in the completion of the group's tasks, reflecting the characteristics of joint negotiation.

In regulation focus, the three groups all appear most frequently in "monitoring" behavior ("content monitoring" and "process monitoring"). "Group 2" and "Group 3" have the same conclusion. "Content monitoring" behavior is the most common among them. Learners supervise each other so as to complete the task requirements in time. And in the aspect of "process monitoring" behavior, learners will seek peer help when they have doubts about relevant knowledge when learning materials gave by teachers. However, "Group 5" is different from the other two groups in "monitoring" behavior. The "Group 5" appears more "process monitoring" behavior, followed by "content monitoring" behavior. In terms of task understanding, "Organization" and "emotion", the three groups reached the same conclusion.

The analysis results of "regulation focus" show that social regulation mainly focuses on "content monitoring", "process monitoring" and "task understanding" in collaborative learning. Studies have shown that the regulation of task knowledge is considered to be a high level of regulatory focus, and it plays a key role in promoting successful collaboration and effective learning (Rogat and Linnenbrink-Garcia 2011). Su et al. (2018) found that there are more "monitoring" behaviors to task knowledge in groups with higher scores in collaborative learning, which indicates that the more content monitoring, the more conducive to the construction of knowledge and the improvement of collaborative learning performance of groups. The monitoring of the task process and the understanding of the task requirements are helpful for learners to reach a consensus on the task understanding, reach a consensus on the task objectives, and complete the learning tasks on time (Malmberg et al. 2017). The “process monitoring” behavior is the most in “Group 5”. This conclusion is consistent with the study by Su et al. (2018) Good collaborative learning requires constant monitoring of the learning process. In view of the inconsistency between the conclusions of the “group 5” and the other two groups, we speculate as follows: First, through the discussion records of the three groups of learners, we can find that the “Group 5” will specify a learner to periodically view the time progress and word number requirements of the awareness module, while the other two groups often have “process Monitoring” behavior when they start collaborative editing. Second, in the activity design of Su et al. (2018), learners have obvious role assignments, and they use wiki editing collaboration and do not provide group awareness module for monitoring. And they need learners to constantly remind members to complete tasks after class. There are fewer “organization” behaviors in the three groups, and our analysis may be due to the more detailed collaborative division of the learners in the task planning module. Emotion regulation plays an important role in collaborative learning environment (Jones and Issroff 2005; Zheng and Huang 2016). However, in our research, “emotion” behaviors are the least. Our analysis may be caused by the following reasons. First, our tool was not embedded with the expression package when it was designed so that the learners could not express their views “friendly”. Second, it is also possible that when we design activities, the time required for the learner to complete the task is not enough. In later research, we should pay more attention to the emotional regulation behavior of learners.

5.2 Acceptance of Scaffolding Tools Supported by Social Regulation of Learning

In order to further test the application effect of the tool, this study used the learning analysis tool technology acceptance LAAM questionnaire (Ali et al. 2012) to test the technical acceptance of social regulation of scaffolding tools. The questionnaire uses the Likert five-point scale to test the acceptance of tool from four dimensions: usefulness, ease of use, satisfaction, and perception of the tool. A total of 24 questionnaires were distributed in this study, and 24 questionnaires were collected, of which 24 were valid questionnaires, and the effective rate was 100%. The results of the survey are shown in Table 4.

Table 4. Characteristics of social regulation behavior.

Dimension	Sample size	Mean value	Standard deviation	High score ratio
Usefulness	24	4.38	0.59	79.17%
Ease of use	24	4.08	0.67	75.00%
Satisfaction	24	3.94	0.76	62.50%
Perception	24	4.35	0.47	83.33%

According to the results of the questionnaire, the average score of the scaffolding tools in the four dimensions exceeds the median value of 3, indicating that the learner's overall acceptance of the tool is higher. Among them, the "usefulness" dimension has the highest score, and the high score ratio is 79.17%, indicating that the learner recognizes the tool's support for social regulation of learning in the process of collaboration. The average scores of "ease of use" and "Perception" are both above 4 points, indicating that the learner as a whole think that the tool is relatively simple and "easy to use", the information provided is clear and easy to understand and is expected to continue to be used in future collaborative learning. The score on the "satisfaction" dimension is 3.94, and the high score ratio is 62.50%, indicating that most of the learners are satisfied with the tool.

6 Conclusion

Social regulation of learning is the key ability to ensure the quality and efficiency of collaborative learning. The scaffolds supported by social regulation of learning can provide learners with shared communication space, present cognitive, social and other group awareness information and support evaluation feedback among learners in collaborative situations, which is of great significance for enhancing learners' social regulation of learning level and enhancing collaborative learning effect.

This research enriches the scaffolds supported by the social regulation of learning and is conducive to revealing the positive effects of scaffolding tools on learners' online collaborative learning. However, this study only conducted a preliminary study on the application of scaffolding tools and failed to pay attention to the continuous effect of learners' online social regulation of learning ability improvement. In future research, researchers need to apply and observe more long-term, in-depth and multi-scenarios on the use of scaffolding tools, and examine the persistence and migration of learners' online social regulation of learning ability and the impact on students' learning.

Acknowledgements. This study is supported by the National Nature Science Foundation of China (Grant No. 61877003) and the Humanities and Social Sciences Fund of Chinese Ministry of Education (Grant No. 18YJC740084).

References

Ali, L., Hatala, M., Gasevic, D., Jovanovic, J.: A qualitative evaluation of evolution of a learning analytics tool. Comput. Educ. **58**(1), 470–489 (2012)

Chang, C.J., Chang, M.H., Chiu, B.C., et al.: An analysis of student collaborative problem solving activities mediated by collaborative simulations. Comput. Educ. **114**, 222–235 (2017)

Hadwin, A.F., Bakhtiar, A., Miller, M.: Challenges in online collaboration: effects of scripting shared task perceptions. Int. J. Comput. Support. Collaborative Learn. **13**(3), 301–329 (2018). https://doi.org/10.1007/s11412-018-9279-9

Hadwin, A., Oshige, M.: Self-regulation, coregulation, and socially shared regulation: exploring perspectives of social in self-regulated learning theory. Teach. Coll. Rec. **113**(2), 240–264 (2011)

Hadwin, A.F., Oshige, M., Gress, C.L.Z., Winne, P.H.: Innovative ways for using gStudy to orchestrate and research social aspects of self-regulated learning. Comput. Hum. Behav. **26**(5), 794–805 (2010)
Janssen, J., Bodemer, D.: Coordinated computer-supported collaborative learning: awareness and awareness tools. Educ. Psychol. **48**(1), 40–55 (2013)
Järvelä, S., Hadwin, A.F.: New frontiers: regulating learning in CSCL. Educ. Psychol. **48**(1), 25–39 (2013)
Järvelä, S., et al.: Socially shared regulation of learning in CSCL: understanding and prompting individual- and group-level shared regulatory activities. Int. J. Comput. Support. Collaborative Learn. **11**(3), 263–280 (2016). https://doi.org/10.1007/s11412-016-9238-2
Järvelä, S., Kirschner, P.A., Panadero, E., Malmberg, J., Phielix, C., Jaspers, J., et al.: Enhancing socially shared regulation in collaborative learning groups: designing for CSCL regulation tools. Educ. Tech. Res. Dev. **63**(1), 125–142 (2015)
Jones, A., Issroff, K.: Learning technologies: affective and social issues in computer-supported collaborative learning. Comput. Educ. **44**(4), 0–408 (2005)
Kim, D., & Lim, C.: Promoting socially shared metacognitive regulation in collaborative project-based learning: a framework for the design of structured guidance. Teach. High. Educ. **23**, 1–18 (2017)
Lai, C.-L., Hwang, G.-J., Tu, Y.-H.: The effects of computer-supported self-regulation in science inquiry on learning outcomes, learning processes, and self-efficacy. Educ. Tech. Res. Dev. **66**(4), 863–892 (2018). https://doi.org/10.1007/s11423-018-9585-y
Laru, J., Malmberg, J., Jarvenoja, H., et al.: Designing simple tools for socially shared regulation: experiences of using Google docs and mobile SRL tools in mathematics education. In: Computer Supported Collaborative Learning (2015)
Lin, J.W., Tsai, C.W., Hsu, C.C., Chang, L.C.: Peer assessment with group awareness tools and effects on project-based learning. Interact. Learn. Environ. **0**(0), 1–17 (2019)
Malmberg, J., Järvelä, S., Järvenoja, H.: Capturing temporal and sequential patterns of self-, co-, and socially shared regulation in the context of collaborative learning. Contemp. Educ. Psychol. **49**, 160–174 (2017)
Miller, M., Hadwin, A.: Scripting and awareness tools for regulating collaborative learning: changing the landscape of support in CSCL. Comput. Hum. Behav. **52**, 573–588 (2015)
Rogat, T.K., Linnenbrink-Garcia, L.: Socially shared regulation in collaborative groups: an analysis of the interplay between quality of social regulation and group processes. Cogn. Instr. **29**(4), 375–415 (2011)
Splichal, J.M., Oshima, J., Oshima, R.: Regulation of collaboration in project-based learning mediated by CSCL scripting reflection. Comput. Educ. **125**(August 2017), 132–145 (2018)
Su, Y., Li, Y., Hu, H., Rosé, C.P.: Exploring college English language learners' self and social regulation of learning during wiki-supported collaborative reading activities. Int. J. Comput. Support. Collaborative Learn. **13**(1), 35–60 (2018). https://doi.org/10.1007/s11412-018-9269-y
Zheng, L., Huang, R.: The effects of sentiments and co-regulation on group performance in computer supported collaborative learning. Internet High. Educ. **28**, 59–67 (2016)
Zheng, Y., Xu, C., Li, Y., Su, Y.: Measuring and Visualizing Group Knowledge Elaboration in Online Collaborative Discussions. J. Educ. Technol. Soc. **21**, 91–103 (2018)
Zimmerman, B.J., Schunk, D.: Handbook of Self-Regulation of Learning. Routledge, Abingdon (2011)

Facilitating Course Recommendations by Word2vec Paradigm Through Social Tags

Jingjing Wang[1], Haoran Xie[2], Oliver Tat Sheung Au[1], Lap-Kei Lee[1], Di Zou[3(✉)], and Fu Lee Wang[1]

[1] School of Science and Technology, The Open University of Hong Kong, Ho Man Tin, Kowloon, Hong Kong
[2] Department of Computing and Decision Sciences, Lingnan University, Tuen Mun, New Territories, Hong Kong
[3] Department of English Language Education, The Education University of Hong Kong, Tai Po, New Territories, Hong Kong
dzou@eduhk.hk

Abstract. As an important educational form, online learning has attracted millions of registered learners, and a huge number of courses are available online. However, it is challenging for learners to identify appropriate courses from a large course pool due to the difficulties of mapping complex learning needs to the high-level course semantics. Several studies in the field of Natural Language Processing (NLP) have recently gained promising performance in capturing the semantic information. In this study, we use these NLP techniques to understand the semantics of learning needs and courses. Specifically, we model users' historical course records as word sentences using skip-gram with negative sampling to obtain course semantics. Furthermore, we introduce Laplacian Eigenmaps as the objective function and integrate the course social tags and course-user interaction as penalty factors to fine-tune the course vectors, especially the courses of different categories but similar contexts. The result verifies that the proposed method is effective for recommending suitable courses for users.

Keywords: Recommender systems · Word2vec · Word embedding · Skip-gram technique

1 Introduction

With the rapid development of the Internet and Web technologies, online learning has attracted millions of registered students worldwide (Henderikx et al. 2017). There are several reasons people tend to choose online learning. For example, learning online is cost-efficient. Another reason is that there are no restrictions on time and location for learning. Some universities or institutions also offer high-quality education free to anyone in the online learning platforms like MOOCs. The learning environment is diverse for learners worldwide (Castle and McGuire 2010). Besides, learning resources are multimedia objects, such as videos, animations, and images (Lau et al. 2013). However,

L.-K. Lee et al. (Eds.): ICTE 2020, CCIS 1302, pp. 239–247, 2020.
https://doi.org/10.1007/978-981-33-4594-2_20

confronted with a growing number of courses, it is challenging to pick up a satisfying course from the large course pool in an online learning platform as the semantics of complex learning needs and courses are difficult to understand and match. A search engine is one of the solutions for addressing this problem. In this method, course descriptions and categories are the main factors for displaying the search results. In other words, if the learner's need is obvious and can be specified by some keywords, the search engine will be sufficient for learners to identify their preferred courses. However, learners may not understand their learning needs very clear in most cases. Furthermore, online courses usually have limited text descriptions, which can be used as keyword search indices. These keywords cannot fully represent the high-level semantics of the courses. Therefore, the search engine is not a good solution when the learning needs are vague, or the courses' indices are insufficient.

The recommender systems (Resnick and Varian 1997) have been another effective solution for the above problem. From e-commerce (e.g., Amazon (Linden et al. 2003) or Alibaba) to online services (e.g., hotels (Zhang et al. 2015) or movie (Bennett and Lanning 2007)), recommender systems can assist users in clarifying their true needs and decision-making in every daily life. Recommender systems attempt to establish the linkages between users and items based on user interaction and predict users' preferences. Collaborative Filtering (CF) was one of the most popular techniques in recommender systems (Herlocker et al. 2004). In a CF-based recommender system, similar items (users) will be recommended based on the user's historical preferences. Although CF has achieved an excellent performance of recommendation, there are still some problems in this approach. A well-known issue of the CF-based approach is the cold-start problem. The system cannot understand a new user's preference as there is no or little interaction between the system and the new user.

Moreover, current course recommender systems rarely consider course contextual information in the historical user records (Ma et al. 2017; Bridges et al. 2018; Hidasi et al. 2018). The course position in the records contains semantic information and can be clustered by the semantic information. Relevant courses will share a common context in the historical user records during a period. Word2vec (Mikolov et al. 2013a, b) is one of the most commonly used Natural Language Processing (NLP) techniques that transform the unstructured natural language to the normalized and structured data with semantic information. Another issue is that learning is a spiral process of knowledge acquisition (Diamond et al. 2008). If someone tries to gain a skill, she will choose some similar courses at the first and most light-similar courses in the second step. Usually, similar courses are in the same category and have similar social tags, while the light-similar courses are not. It should be noted that the first step and the second step is not independent but alternation. However, Word2vec assumed that words share a common context will have a similar word vector. This hypothesis makes it difficult to distinguish two courses that share the same context but in different categories.

In this paper, we propose a course recommendation method by Word2vec paradigm through social tags. Firstly, we treated historical user records as the training data and used a content-based recommendation method to overcome the cold-start problem. Then we introduce Laplacian Eigenmaps as the objective function, and the course social tags and course-user interaction as the penalty factor, to fine-tune the vectors generated by the language model Word2vec. One of the advantages of the proposed method is that it can still have a stable performance for new users who have limited interactive data. The other advantage is that using penalty factors will weaken the influence of light-similar semantics and improve the recommendation's accuracy.

The remaining sections of this paper are organized as follows: Sect. 2 formalizes the problem of course recommendations. In Sect. 3, we introduce the proposed model which exploits Course2vec and integrates social tags in our proposed course recommender system. The experimental results are shown in Sect. 4. Finally, we summarize this research and discuss future directions in Sect. 5.

2 Problem Formalization

We formalize recommending courses based on historical user records and course attributes into an embedding problem. The goal is to learn $X_C \in \mathbb{R}^{|V| \times d}$, where X_C is a low-dimension vector of courses, and d is a small number of latent dimensions. For each user $u \in \mathrm{U}$, $\mathrm{U} = \left\{u_1, \cdots, u_{|U|}\right\}$, given his profile, i.e., the historical course records $Q_u = (c_1, \cdots, c_n)$ with $c_n \in C$, Q is the set of all user historical course records $Q = (Q_1, \cdots, Q_u)$ and course attribute $C = \left\{T_1, \cdots, T_{|C|}\right\}$, T_i is the course social tag. We aim to obtain the course vectors with historical user course records Q and course attribute C as formula 1.

$$X_C = f(Q, C) \tag{1}$$

3 The Proposed Model

This section gives an overview of the proposed model, then introduces a language algorithm to transform historical user records, and finally uses Laplacian Eigenmaps as the objective function to train the entire model. Our study's core contribution is that we further divide course contextual information into two types of semantic information based on the principle of knowledge acquisition. We define these two types of semantic information as follows:

Similar semantics: Courses share the same contextual information in the historical user records, and in the same category, share common social tags.

Light-similar semantics: Courses share the same contextual information in the historical user records but in different categories and have different social tags.

These two kinds of semantic information will cross-over be occurred in the user learning records. Therefore, Word2vec paradigm's significance through social tags in a recommender system include the following two aspects. Firstly, courses that have the same contextual information will locate closely in the low dimension vectors by extracting the semantic information from user behavior records. Secondly, the course vectors with social tags to distinguish similar semantics and light-similar semantics can be fine-tuned.

3.1 Generic Framework

We propose a course recommender system which exploits Word2vec to obtain semantic information (e.g., Bengio et al. 2013) from historical user records and integrate the course social tags and course-user interaction. Figure 1 illustrates the generic framework of the proposed model. The left part of Fig. 1 denotes using Word2vec to obtain the course vector, while the right one demonstrates integrating social tags to adjust the vectors.

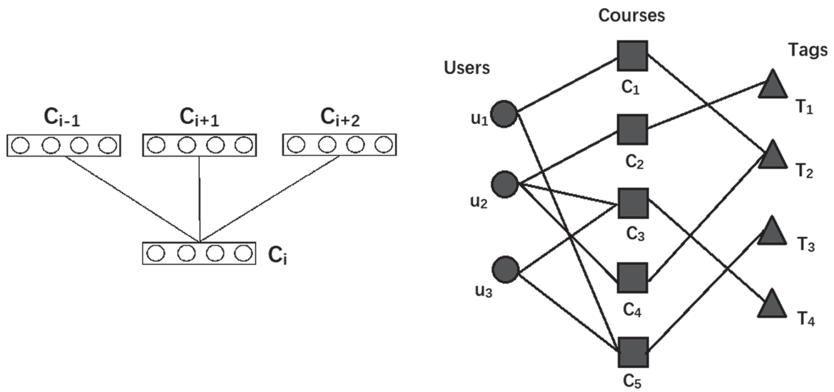


Fig. 1. The overall framework of the proposed model

3.2 Course2vec

Word2vec is most commonly used in NLP. It is a two-layer neural network model of natural language processing topics introduced by Google in 2013 (Mikolov et al. 2015). In Word2vec, sentences are regarded as an ordered sequence of words as the model input, train the network model and generate the vector to represent the word based on the assumption that words share the same context should have a close position in the target dimension. Compared to other language models like TF-IDF and Latent Semantic Analysis, these generate vectors to make the semantic information abundant (Zhao and Shang 2010) and can easily be used in the downstream tasks as word clustering and classification. User historical records can be modeled as sentences in Word2vec because they are both generated by basic elements based on semantic rules. Words co-occurred in the sentence contain semantic information, while relevant courses are located closely in the records during a period.

Word2vec contains two sub-models which are the CBOW model and the Skip-gram model. These two models can be used in different scenarios. In the CBOW model, surrounding words are the input; the model estimates the center words' likelihood. While in the Skip-gram model, center words are the input, the model aims to predict its neighbors. As our goal is to predict the next selected item based on the existing records, this study employs the Skip-gram model.

Similar to Word2vec, we propose the Course2vec to model the course representation. Given the set of courses $Q_u = (c_1, \cdots, c_n)$ taken by one user. Sequences

$Q = (Q_1, \cdots, Q_u)$ are the ordered course list. We aim to maximize the probability of the contextual words in the sequence.

$$y_t = \frac{1}{T}\sum\nolimits_{t=1}^{T}\left(\sum\nolimits_{-c\le j\le c, j\ne 0}\log p\left(c_{t+j}|c_t\right)\right) \tag{2}$$

where $c_1, \ldots, c_T$ are courses in the training corpus, and c is the length of the window around target course c_t. Train the network by feeding it word pairs $<c_{t+j}, c_t>$. The probability $\Pr\left(c_{t+j}|c_t\right)$, the important part of the objective y_t, is given by SoftMax:

$$\Pr\left(c_{t+j}|c_t\right) = \frac{\sum e^{v_{c_{t+j}}.v_{c_t}}}{\sum_{c'\in C} e^{v'_c.v_{c_t}}} \tag{3}$$

where v_{c_t} and $v_{c_{t+j}} \in \mathbb{R}^d$ are vector representations for course c_t and c_{t+j}, respectively. Training the whole neural network indicates maximizing the function $\Pr\left(c_{t+j}|c_t\right)$. In other words, each training $\Pr\left(c_{t+j}|c_t\right)$ will adjust all the neural network cells, which will slow down the training process. Negative sampling addresses this problem by only modify a small number of negative courses to update the weights. The objective function can be defined as follows:

$$J = log\sigma(c_w.c_t) + \sum\nolimits_{i=1}^{k} log\sigma\left(-c_{N,i}.c_t\right) \tag{4}$$

where σ denotes the sigmoid function, and k is the total number of randomly selected negative courses.

3.3 Laplacian Eigenmaps

Although the users' historical records contain contextual information, language models can transform contextual information into vectors. There are two types of semantic information: *similar semantics* and the other is the *light*-s*imilar semantics*. To deal with this issue, we propose to use Laplacian Eigenmaps as the objective function to generate the low-dimensional representation and fine-tune the course vectors obtained by Course2vec, especially courses share common context but in different categories. The objective function is defined as follows:

$$\mathcal{L} = \mu_{ij}\left|\left|y_i - y_j\right|\right|_2^2 \tag{5}$$

where y_i, y_j is the final vector of courses i, j, μ_{ij} is the weight between them. In the model, μ_{ij} is a penalty factor, if there is a slight penalty, course i and course j will locate far in the final space. We aim to minimize the objective function to ensure that y_i and y_j are close if course i and course j have a heavy penalty.

$$argmin\,\mathcal{L} = \mu_{ij}\left|\left|y_i - y_j\right|\right|_2^2 \tag{6}$$

In our model, there are two factors related to μ_{ij}. The first one is the number of common interactive users, and the second one is the number of social tags, μ_{ij} can be defined as follows.

$$\mu_{ij} = \frac{U_i \cap U_j}{U_i \cup U_j} + \frac{T_i \cap T_j}{T_i \cup T_j} \tag{7}$$

where U_i is the set of user enroll course i and U_j is the set of user enroll course j, T_i is the set of social tags of the course i, and T_j is the set of course j social tags. These social tags are obtained by using Text-CNN (Kim 2014) from analyzing the course introduction and user online comments. To solve our model's objective function, we regard the course as the node in the graph. Given a network $G = (V, E)$, in which V and E are the set of course node and edge, respectively. We can obtain course adjacency matrix S. For each instance in S, $w_i = \left\{\mu_{ij}\right\}_{j=1}^{n}$, $\mu_{i,j}$ means the links between c_i and c_j. Therefore, the objective function can be rephrased as follows:

$$argmin\,\mathcal{L} = \mu_{ij}\left|\left|y_i - y_j\right|\right|_2^2 = 2tr\left(Y^T LY\right) \tag{8}$$

where $L = D - S$, D is the diagonal matrix, and $D_{i,i} = \sum_j s_{i,j}$, Y is the prediction vector. Finally, the model generates the final representation of the course with the user's historical behavior information. In this paper, cosine similarity is employed for measuring the courses' similarity with the low-dimensional vectors. At last, *top-N* most similar courses will be recommended to the user based on the course similarity.

4 Experiment

In this paper, the dataset was introduced in the section of MOOC data selected from XuetangX. XuetangX, launched in October 2012, provided over 1,000 courses distributed 12 categories from art, biology to computer science. More than 10,000,000 users have registered in XuetangX. In this experiment, we selected enrolled behaviors from October 1st, 2016 to December 30th, 2017, as samples. Each instance in the training set was a sequence of historical records. For each course, we constructed social tags by using text-CNN to extract from the course introduction. In this experiment, we selected the *top-5* social tags as the course tags. Table 1 showed the social tags of some courses.

Table 1. The social tags of sample courses

Course	*Social tags*
Java Programming: Solving Problems with Software	Programming, Duke university, Problem-solving skills, Java program, Harder
Introduction to Computer Science and Programming	Computer Graphics, Mathematic, JavaScript, Computer Programming, Beginner
Mathematics for Machine Learning	Machine Learning, Beginner, Data Science, Python Programming, Gradient Descent

Firstly, we sorted the user-selected historical records during the period as the dataset. Then, we built a Course2vec model using python and genism to transform the semantic information to the word representation. Finally, to avoid courses in different categories share common context in Course2vec, we used the Laplacian Eigenmaps method as the objective function to fun-tune the Course2vec vectors. The penalty factor in the objective function was based on user-item-tag interaction.

Metrics and Baselines. We compared our proposed method with the following two methods measured by the two metrics: Recall@20 and Precision@20.

Collaborative Filtering (CF): This method regarded the user-item interactions as the original user vector. If the user enrolled in the course, the interactive matrix's value equals 1, else 0.

Matrix Factorization (MF): This method modeled user preferences by decomposing the user-item interactive matrix to obtain user and item embeddings.

Parameter Settings. To gain the course vectors by Course2vec, we constructed the model with a 64-dimensional course vector as output, selecting five neighbors as the context, five courses as the negative, and neglecting the course that had been occurred less than five times. Here we chose a relative high dimension vector because high dimension reserves rich information, and in the following step, Laplacian Eigenmaps reduced the number of vector dimensions to 32. According to each course's selected records, we obtained the similarity of each course by the course vector and recommend *top-N* courses.

Table 2. Performance of course recommendations by using different models

Method	*Recall@20*	*Precision@20*
CF	32.10%	21.96%
MF	35.30%	22.50%
Course2vec	35.70%	22.70%
Course2vec ++	35.80%	23.00%

Table 2 showed the results of course recommendations in terms of Recall@20 and Precision@20. Course2vec ++ was a model of the Course2vec with course social tags and course-user interactions. The proposed Course2vec and Course2vec ++ models outperformed all baselines. The second best performance achieved by Course2vec also verified that incorporating semantic information can improve the effectiveness of the recommendation. Course2vec ++ achieved the best performance and was lightly better than Course2vec. Table 3 showed a qualitative example of recommendation results for a specific user (i.e., the user id is 58 in the dataset) by using two different models (i.e., Course2vec and Course2vec ++).

Table 3. A qualitative example of recommendation results of a specific user (id 58)

Course name	*Recommend Results*	
	Course2vec	Course2vec ++
Advanced Machine Learning	Mathematics for Machine Learning	Mathematics for Machine Learning
	Machine Learning with TensorFlow	Machine Learning with TensorFlow
	Python 3 Programming	Data Science: Statistics and Machine Learning
	Data Science: Statistics and Machine Learning	Deep Learning
	Python for Data Science and AI	Machine Learning with Python

5 Conclusion

This paper proposed a method of course recommendation. First, we represented the user's historical records as word sentences through exploiting the skip-gram with negative sampling to obtain course embeddings. To avoid two types of semantic information that share common context in the records, we integrated the course social tags and course-user interactions as penalty factors to adjust course embeddings under different categories. Experimental results showed that Course2vec captured semantic information from the records. Compared to recommendations based on Course2vec, Course2vec + + improved the recommendation accuracy. However, this paper focused on observed interactions between users and courses from the perspective of their bipartite graph. For our future research, we plan to integrate information extracted from unobserved items to improve the recommendations' accuracy and serendipity.

Acknowledgement. The work described in this paper was supported by the Katie Shu Sui Pui Charitable Trust — Academic Publication Fellowship (Project Reference No. KSPF2019-03).

References

Bengio, Y., Courville, A., Vincent, P.: Representation learning: a review and new perspectives. IEEE Trans. Pattern Anal. Mach. Intell. **35**(8), 1798–1828 (2013)

Castle, S.R., McGuire, C.J.: An analysis of student self-assessment of online, blended, and face-to-face learning environments: Implications for sustainable education delivery. Int. Educ. Stud. **3**(3), 36–40 (2010)

Diamond, N., Koernig, S.K., Iqbal, Z.: Uniting active and deep learning to teach problem-solving skills: Strategic tools and the learning spiral. J. Mark. Educ. **30**(2), 116–129 (2008)

Henderikx, M.A., Kreijns, K., Kalz, M.: Refining success and dropout in massive open online courses based on the intention–behavior gap. Distance Educ. **38**(3), 353–368 (2017)

Herlocker, J.L., Konstan, J.A., Terveen, L.G., Riedl, J.T.: Evaluating collaborative filtering recommender systems. ACM Trans. Inf. Syst. (TOIS) **22**(1), 5–53 (2004)

Kim, Y.: Convolutional neural networks for sentence classification. arXiv preprint arXiv:1408.5882 (2014)

Lau, R.W.H., Yen, N.Y., Li, F., Wah, B.: Recent development in multimedia e-learning technologies. World Wide Web **17**(2), 189–198 (2013). https://doi.org/10.1007/s11280-013-0206-8

Linden, G., Smith, B., York, J.: Amazon. com recommendations: Item-to-item collaborative filtering. IEEE Internet Comput. **7**(1), 76–80 (2003)

Mikolov, T., Chen, K., Corrado, G., Dean, J.: Efficient estimation of word representations in vector space. arXiv preprint arXiv:1301.3781 (2013a)

Mikolov, T., Sutskever, I., Chen, K., Corrado, G.S., Dean, J.: Distributed representations of words and phrases and their compositionality. In: Advances in neural information processing systems, pp. 3111–3119 (2013b)

Mikolov, T., Chen, K., Corrado, G.S., Dean, J.A.: U.S. Patent No. 9,037,464. U.S. Patent and Trademark Office, Washington, DC (2015)

Resnick, P., Varian, H.R.: Recommender systems. Commun. ACM **40**(3), 56–58 (1997)

Bridges, C., Jared, J., Weissmann, J., Montanez-Garay, A., Spencer, J., Brinton, C.G.: Course recommendation as graphical analysis. In: 2018 52nd Annual Conference on Information Sciences and Systems (CISS), pp. 1–6. IEEE (March 2018)

Bennett, J., Lanning, S.: The netflix prize. In: Proceedings of KDD cup and workshop, vol. 2007, p. 35 (August 2007)

Zhao, Z.D., Shang, M.S.: User-based collaborative-filtering recommendation algorithms on hadoop. In: 2010 Third International Conference on Knowledge Discovery and Data Mining, pp. 478–481. IEEE (January 2010)

Zhang, K., Wang, K., Wang, X., Jin, C., Zhou, A.: Hotel recommendation based on user preference analysis. In: 2015 31st IEEE International Conference on Data Engineering Workshops, pp. 134–138. IEEE (April 2015)

Ma, H., Wang, X., Hou, J., Lu, Y.: Course recommendation based on semantic similarity analysis. In: 2017 3rd IEEE International Conference on Control Science and Systems Engineering (ICCSSE), pp. 638–641. IEEE (August 2017)

Hidasi, B., Karatzoglou, A.: Recurrent neural networks with top-k gains for session-based recommendations. In: Proceedings of the 27th ACM International Conference on Information and Knowledge Management, pp. 843–852 (October 2018)

Social Media in Higher Education: A Review of Their Uses, Benefits and Limitations

Kam Cheong Li and Billy Tak-Ming Wong(✉)

The Open University of Hong Kong, Ho Man Tin, Hong Kong
{kcli,tamiwong}@ouhk.edu.hk

Abstract. As a prevalent type of internet technology for social interaction and collaboration, social media have been increasingly used for educational purposes. This paper presents a review of the research work on their uses, benefits and limitations for teaching and learning in higher education contexts. It covers a total of 77 articles published from 2010 to 2019 which were collected from Scopus and Google Scholar. The uses of social media revolved around two major areas—as a learning management system and a means to enhance students' engagement. The benefits of using social media covered students' satisfying experience in collaborative learning, engagement in learning activities, real-time communication, and knowledge sharing, as well as communities of practice; and such benefits have been shown to have a relationship with the improvement of learning outcomes. The limitations covered concerns about blurring boundaries between public and private life and between teachers and students; the effectiveness for the construction of knowledge; a lack of pedagogical and technical skills; and compatibility with instructional design. The studies also reveal the changing role of teachers and their need to innovate teaching practices with effectively utilised features of social media. Finally, the results suggest future work to be done for promoting the educational use of social media and enhancing their effectiveness, such as support from institutional policies on technology-enhanced learning and the promotion of teaching innovations.

Keywords: Social media · Web 2.0 · Higher education · Engagement · Interaction · Learning outcomes

1 Introduction

Social media, also called Web 2.0 and social software, are the types of internet technology that provide a channel for social interaction and collaboration (Dabbagh and Kitsantas [7]). They allow users to engage in and share content in a digital environment. The use of social media has become increasingly prevalent. For example, the Pew Research Centre [32] found that, in 2019, more than 70% of adults in the United States used at least one type of social media, and 90% of those aged between 18 and 29 did so.

Along with the popularity of social media, there has been a growing interest in its use for teaching and learning purposes. For example, social media have been used by educational institutions as a kind of learning management system (LMS). Skype has

L.-K. Lee et al. (Eds.): ICTE 2020, CCIS 1302, pp. 248–257, 2020.
https://doi.org/10.1007/978-981-33-4594-2_21

been employed to allow students to join first-hand interviews with domain experts; and WordPress has been adopted by students to develop their e-portfolios for performance-based assessment (Dabbagh and Reo [8]).

The various initiatives on social media have demonstrated its potential to support active and informal learning so as to strengthen students' interaction and engagement inside and outside classes. A key feature of social media is the production of user-generated content. Rennie and Morrison [35] claimed that promoting learners' generation of educational content with the aid of social media enables them to actively construct their knowledge instead of passively absorbing information provided by teachers. More importantly, the virtual community space on social media supports learners' communication with others with similar learning preferences from around the world, which motivates them to learn and persist.

However, there are also limitations in the educational use of social media. For instance, it is challenging to strike a balance between the informal and academic use of social media. Manca and Ranierit [22] found that the pedagogical affordances of Facebook have not been fully realised and obstacles such as cultural issues can deter people from adopting it as an effective learning environment. How to overcome the barriers in order to realise the pedagogical potential of social media remains an important issue to be addressed.

This paper aims to give a comprehensive and systemic review of the research work on this area in order to provide an overview of the use of social media in higher education. It focuses on the ways social media have been used in teaching and learning, as well as the benefits and limitations.

2 Related Studies

The increasing significance of social media—particularly among young people—has resulted in the proliferation of related studies. For example, Balakrishnan and Gan [4] analysed its usage for learning in relation to university students' learning styles. They revealed a range of factors influencing students' intention to use social media for learning, such as their self-efficacy and perceived usefulness, as well as the communication functionalities of social media. Also, Manca and Ranieri [23, 24] examined the patterns of higher education scholars on using social media in academic practices, and identified the key variables affecting the frequency of use, such as the scientific discipline, age and seniority of the scholars.

As regards the benefits of social media use, Hung and Yuen [14] found that students could develop social connectedness and increase engagement when social networking sites were used as supplementary tools for learning. In addition, Gikas and Grant [13] traced students' use of social media together with mobile devices in the courses of three universities for two semesters, and showed that the students demonstrated enhanced collaboration and interaction, as well as engagement in content creation, when they used social media.

However, the relevant literature has also raised problems when implementing social media in teaching and learning. For instance, Brooks and Pomerantz [6] surveyed students' feedback on social media use and showed that nearly half of them preferred less use of social media by instructors as a learning tool. They pointed out that students may be less enthusiastic about engaging in learning activities involving social media which require them to "use their private devices and accounts or produce content for the consumption of others", "unless faculty clearly and openly state the pedagogical benefits … to enhance content, context, and learning expectations" (p. 24). Also, Moran, Seaman and Tinti-Kane [29] presented the barriers to widespread adoption of social media for teaching as perceived by faculty staff where the top concerns lay in privacy issues and the integrity of student submissions.

The literature reviews in this area, however, have not yet presented a comprehensive summary of relevant findings. They have only provided an overview of how different researchers have approached the topic, or their general views on the educational potential of social media (Evans [11]; Manca and Ranierit [22]). This paper aims to address the literature gap through analysing the academic practices on social media.

3 Methodology

This study aims to investigate the case studies of higher education institutions involving teaching and learning practices on social media, as well as the benefits and limitations as revealed from the practices. Relevant cases studies were collected from Scopus and Google Scholar, using the key terms ("social media," "higher education" and "teaching"). The period of the publications was 2010 to 2019. The articles collected were further checked by filtering out the duplicated ones, as well as those not written in English. In addition, the following inclusion criteria were applied for selecting relevant case studies:

i. They reported one or more empirical practice involving the use of social media in a higher education institution.
ii. They contained the details of social media use, such as the purpose, detailed description of implementation, outcomes, and evaluation.

A total of 77 case studies were selected following the above criteria. They were analysed in terms of the characteristics of the social media use. The benefits and limitations were also identified and categorised from the case studies according to the outcomes or evaluation results reported.

4 Results

4.1 Use of Social Media

The use of social media as reported in the case studies revolved around two major areas—as a learning management system and a means to enhance students' engagement.

As a Learning Management System
Some institutions used social media as an alternative to learning management systems (LMS) such as Moodle or Blackboard, with the aim of enhancing the involvement of students outside the class. The social media sites served as a platform to display

course announcements, manage course materials, and collect assignments. Their use was intended to address the unfavourable attitude of students towards the design of traditional LMSs, which were then used mainly in campuses as a consequence (Deng and Tavares [10]). In contrast, the community-building and sharing functions of social media would be more attractive for students and motivate them to access the social media sites outside the campus. For example, Albayrak and Yildirim [2] reported the practice in a university in Turkey where instructors made use of Facebook to share course information, make announcements, provide learning activities, and conduct discussion.

As a Means to Enhance Students' Engagement
It is also common for institutions to make use of social media to conduct learning activities as a means to enhance students' engagement in both formal and informal contexts. The activities may cover those conducted in class or students' self-directed course-related activities outside the classroom, such as interaction among peers. For example, Menkhoff, Chay, Benggtsson, Woodard, and Gan [28] used Twitter for students to express views and ask questions during class time. Also, McCarthy [25] reported a practice in the University of Adelaide where students were required to upload their work to Facebook and comment on peers' submissions. In this regard, Irwin, Ball and Desbrow [15] said that social media offer a new space for developing academic connections, communication and collaboration.

4.2 Benefits

Figure 1 shows the benefits of social media use and the percentages of case studies which reported them. Overall, the benefits covered students' experience in collaborative learning, engagement in learning activities, real-time communication, and knowledge sharing, as well as communities of practice.

Fig. 1. Percentages of case studies which report the benefits of social media use

Offering Students a Satisfying Collaborative Learning Experience
Social media are perceived to have the potential to foster collaborative learning. Through its use, students are able to interact with their peers on problem-solving, and co-construct and share knowledge. Given the more satisfying learning experience, they may also pay more attention in the class and be eager to discuss course-related issues after class.

This benefit has been most commonly reported in the case studies. For example, Laborda and Litzler [17] presented students' feedback on Web 2.0 tools, with 90% of them agreeing that the tools could promote cooperation among classmates.

Attracting Students to Visit Them and Engage in Learning Activities
Compared with traditional LMSs, students have tended to visit social media more frequently. This benefit has been widely reported as measured by the numbers of comments, views, and 'likes' for the social media sites. Lieberman [20] showed the popularity of Facebook among students—it was checked by about 60% of students several times a day, while Blackboard was checked only once a day by a similar proportion of students.

The more frequent use of social media can also help to promote students' engagement in learning activities. In general, students were more active in social media in terms of participating in discussion, asking questions, joining activities, and reading others' posts (Hung and Yuen [14]). Also, McCarthy [26] reported students' feedback that in online discussion they felt more comfortable in expressing themselves and commenting on others' work or ideas.

Enabling Two-Way Communication in Real Time
Social media feature real-time communication and serve as channels for close interaction between instructors and students that allows individual mentorship. Relevant case studies have shown students' preference for communicating and discussing course-related materials with fellow students via social media. For example, Ahern, Feller and Nagle [1] reported students' comments that communication on Facebook was efficient because members of Facebook groups shared a relevant background and information. O'Boyle [31] also presented students' positive perception of Facebook and Twitter which allow users to interact and receive responses efficiently. They also promoted face-to-face interactions between students and instructors, as students would feel more comfortable approaching in person instructors who have already interacted with them via the social media (O'Boyle [31]).

Enhancing Knowledge Sharing Among Users
Social media have been shown to be effective platforms for requesting and sharing subject-related materials. Their ease of use and immediate communication are two main factors which make its use attractive for sharing information. Studies such as Neier and Zayer [30] and Sobaih and Moustafa [37] presented students with social media sites to share course materials, and the students reported their positive attitude and experience where they could do so without anxiety.

Fostering Communities of Practice
The enhancement of communication among different learners and teachers on social media sites has promoted a community of inquiry and community of practices, as well as fostering a close relationship among them. For example, Feliz, Ricoy and Feliz [12] found learners' high degree of commitment to Twitter, which helped them to create a sense of belonging to the community.

4.3 Limitations

Figure 2 shows the limitations of social media use and the percentages of case studies which reported them. The limitations cover concerns about blurring boundaries between public and private life and between teachers and students; the effectiveness for the construction of knowledge; a lack of pedagogical and technical skills; and compatibility with instructional design.

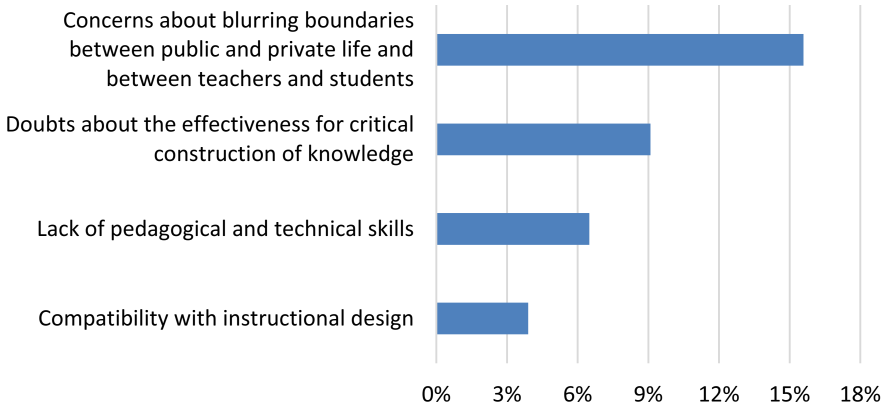


Fig. 2. Percentages of case studies which report the limitations of social media use

Concerns About Blurring Boundaries Between Public and Private Life and Between Teachers and Students
Despite the effectiveness of social media in promoting interaction, students may have reservations about accepting teachers and classmates as their "friends" on social media. Some students hope that their academic discussion is kept private, instead of being viewed by all of their friends on social media. For example, Megele [27] raised the issue of how to balance individual privacy and the ease of disseminating learning materials as an important ethical consideration for institutions.

Also, while the distance between teachers and students has been decreased through their interaction on social media, studies have also pointed out users' concerns about blurring the boundaries between their relationships that would affect the effectiveness of teaching and learning. Shelton [36] surveyed faculty staff's feedback on this and revealed the limitation of Facebook as a social learning space.

Doubts About the Effectiveness for Critical Construction of Knowledge
While social media are effective in promoting students' engagement, their effectiveness in enhancing learning outcomes remains doubtful. For instance, Rambe [33] illustrated the difficulties in facilitating quality academic discussions on social media sites which may foster students' learning at the conceptual level. Some students may just read the messages and share course-related information on social platforms but seldom respond to questions (Lin, Hoffman, and Borengasser [21]). Besides, the length limit of each

message imposed by some social media may also hinder reflective thinking and in-depth discussion (Kassens-Noor [16]).

Lack of Pedagogical and Technical Skills
The effective use of social media also depends on teachers' pedagogical and technological skills. Studies have revealed that some teachers lack sufficient skills and may question the potential pedagogical benefits of social media, or follow the traditional ways of using LMS (Veletsianos, Kimmons, and French [38]). Students may also encounter technical difficulties during the classes that are instructed with social media. For example, Demirbilek [9] found that, despite their familiarity with the general use of social media sites, students may not be familiar with learning activities (e.g. wiki-based learning) carried out on them.

Compatibility with Instructional Design
The use of social media alone may not lead to desirable pedagogical outcomes. Some studies have suggested that the effective pedagogical use of social media should be combined with suitable instructional design, where little integration between them may deter the effectiveness of practices (Rambe and Nel [34]).

5 Discussion and Conclusion

The results of this study show the use of social media sites by higher education institutions for educational purposes. They have been widely adopted as channels to enrich students' learning engagement. The features of social media on community- building and easy sharing of information have been shown to promote learning as an alternative to or replacement of traditional LMS. This suggests the importance of social interaction as an essential part of students' online learning experience, which has been verified in relevant studies (Li, Lee, Wong, Yau, and Wong [18]).

The benefits of social media as shown in the case studies cover mainly the enhancement of visibility and accessibility of course materials as well as learners' participation in online discussion and collaboration. These advantages are highly attributed to the popularity of social media among students—nearly all of them have at least one account and tend to be active on the sites, and so can check and reply to posts frequently. The benefits have also been shown to have a relationship with the improvement of learning outcomes, such as academic performance (Al-rahmi, Othman, Yusof and Musa [3]) and skills (e.g. critical thinking, summarising and collaborating skills) (Demirbilek [9]).

The limitations of social media use reveal the concerns of learners and teachers about adopting this channel for educational purposes, such as blurring the boundaries between public and private life as well as between students and teachers, and the changing role and practice of teachers in the classroom. While teachers' involvement is vital for the educational use of social media, the social media themselves should also be accompanied by a clear pedagogical objective (Brooks and Pomerantz [6]) and existing instructional practices may not be as effective on this channel as those in traditional learning environments (Manca and Ranieri [23, 24]).

Such limitations and concerns suggest future work to be done for promoting the educational use of social media and enhancing their effectiveness. For example, Bennett, Bishop, Dalgarno, Waycott, and Kennedy [5] proposed early the integration of traditional LMS with Web 2.0 tools to extend their functions for improving interaction with students using the ways popular among them. Institutional policies should also include support in technological and pedagogical aspects for students and teachers on the educational use of social media to address the limitations. This would involve the refinement of relevant frameworks for institutional policies on technology-enhanced learning (Wong, Wong and Pang [40]), the promotion of teaching innovations among teachers (Wong [39]), and the exploration of means to incorporate the features of social media into personalised learning approaches (Li and Wong [19]).

References

1. Ahern, L., Feller, J., Nagle, T.: Social media as a support for learning in universities: an empirical study of Facebook groups. J. Decis. Syst. **25**(sup1), 35–49 (2016)
2. Albayrak, D., Yildirim, Z.: Using social networking sites for teaching and learning: students' involvement in and acceptance of Facebook as a course management system. J. Educ. Comput. Res. **52**(2), 155–179 (2015)
3. Al-rahmi, W.M., Othman, M.S., Yusof, L.M., Musa, M.A.: Using social media as a tool for improving academic performance through collaborative learning in Malaysian higher education. Rev. Eur. Stud. **7**(3), 265–275 (2015)
4. Balakrishnan, V., Gan, C.: Students' learning styles and their effects on the use of social media technology for learning. Telematics Inform. **33**(3), 808–821 (2016)
5. Bennett, S., Bishop, A., Dalgarno, B., Waycott, J., Kennedy, G.: Implementing web 2.0 technologies in higher education: a collective case study. Comput. Educ. **59**(2), 524–534 (2012)
6. Brooks, D.C., Pomerantz, J.: ECAR study of undergraduate students and information technology, 2017. ECAR, Colorado (2017)
7. Dabbagh, N., Kitsantas, A.: Personal learning environments, social media, and self-regulated learning: A natural formula for connecting formal and informal learning. Internet High. Educ. **15**(1), 3–8 (2012)
8. Dabbagh, N., Reo, R.: Impact of Web 2.0 on higher education. In: Surry, D.W., Stefurak, T., Gray, R. (eds.) Technology integration in higher education: social and organizational aspects, pp. 174–187. IGI Global, Pennsylvania (2011)
9. Demirbilek, M.: Social media and peer feedback: What do students really think about using Wiki and Facebook as platforms for peer feedback? Act. Learn. High. Educ. **16**(3), 211–224 (2015)
10. Deng, L., Tavares, N.J.: From moodle to Facebook: exploring students' motivation and experiences in online communities. Comput. Educ. **68**, 167–176 (2013)
11. Evans, C.: Twitter for teaching: Can social media be used to enhance the process of learning? Br. J. Educ. Technol. **45**(5), 902–915 (2014)
12. Feliz, T., Ricoy, C., Feliz, S.: Analysis of the use of Twitter as a learning strategy in master's studies. Open Learn. J. Open Distance e-Learn. **28**(3), 201–215 (2013)
13. Gikas, J., Grant, M.: Mobile computing devices in higher education: Student perspectives on learning with cellphones, smartphones & social media. Internet High. Educ. **19**, 18–26 (2013)
14. Hung, H.T., Yuen, S.C.Y.: Educational use of social networking technology in higher education. Teach. High. Educ. **15**(6), 703–714 (2010)
15. Irwin, C., Ball, L., Desbrow, B.: Students' perceptions of using Facebook as an interactive learning resource at university. Australas. J. Educ. Technol. **28**(7), 1221–1232 (2012)

16. Kassens-Noor, E.: Twitter as a teaching practice to enhance active and informal learning in higher education: The case of sustainable tweets. Act. Learn. High. Educ. **13**(1), 9–21 (2012)
17. Laborda, J., Litzler, M.: English for business: student responses to language learning through social networking tools. ESP Today **5**(1), 91–107 (2017)
18. Li, K.C., Lee, L.Y.K., Wong, S.L., Yau, I.S.Y., Wong, B.T.M.: Effects of mobile apps for nursing students: learning motivation, social interaction and study performance. Open Learn. J. Open Distance Learn. **33**(2), 99–114 (2018)
19. Li, K.C., Wong, B.T.M.: How learning has been personalised: a review of literature from 2009 to 2018. In: Cheung, S., et al. (eds.) Blended Learning: Educational Innovation for Personalized Learning. Lecture Notes in Computer Science, vol. 11546, pp. 72–81. Springer, Berlin (2019). https://doi.org/10.1007/978-3-030-21562-0_6
20. Lieberman, S.: Using Facebook as an interactive learning environment in European political studies. Eur. Polit. Sci. **13**(1), 23–31 (2014)
21. Lin, M.F.G., Hoffman, E.S., Borengasser, C.: Is social media too social for class? A case study of twitter use. Tech. Trends **57**(2), 39–45 (2013)
22. Manca, S., Ranierit, M.: Is it a tool suitable for learning? A critical review of the literature on Facebook as a technology-enhanced learning environment. J. Comput. Assist. Learn. **29**, 487–504 (2013)
23. Manca, S., Ranieri, M.: "Yes for sharing, no for teaching!": social media in academic practices. Internet High. Educ. **29**, 63–74 (2016)
24. Manca, S., Ranieri, M.: Facebook and the others. Potentials and obstacles of social media for teaching in higher education. Comput. Educ. **95**, 216–230 (2016)
25. McCarthy, J.: Blended learning environments: using social networking sites to enhance the first year experience. Australas. J. Educ. Technol. **26**(6), 729–740 (2010)
26. McCarthy, J.: Enhancing feedback in higher education: students' attitudes towards online and in-class formative assessment feedback models. Act. Learn. High. Educ. **18**(2), 127–141 (2017)
27. Megele, C.: eABLE: embedding social media in academic curriculum as a learning and assessment strategy to enhance students learning and e-professionalism. Innov. Educ. Teach. Int. **52**(4), 414–425 (2015)
28. Menkhoff, T., Chay, Y.W., Benggtsson, M.L., Woodard, C.J., Gan, B.: Incorporating microblogging ("tweeting") in higher education: lessons learnt in a knowledge management course. Comput. Hum. Behav. **51**, 1295–1302 (2015)
29. Moran, M., Seaman, J., Tinti-Kane, H.: Blogs, Wikis, Podcasts And Facebook: how today' s higher education faculty use social media. Pearson Learning Solutions, Boston (2012)
30. Neier, S., Zayer, L.T.: Students' perceptions and experiences of social media in higher education. J. Mark. Educ. **37**(3), 133–143 (2015)
31. O'Boyle, I.: Mobilising social media in sport management education. J. Hospitality, Leisure, Sport Tourism Educ. **15**, 58–60 (2014)
32. Pew Research Centre: Social Media Fact Sheet (2019). http://www.pewinternet.org/fact-sheet/social-media/
33. Rambe, P.: Constructive disruptions for effective collaborative learning: navigating the affordances of social media for meaningful engagement. Electron. J. e-Learn. **10**(1), 132–146 (2012)
34. Rambe, P., Nel, L.: Technological utopia, dystopia and ambivalence: teaching with social media at a South African university. Br. J. Edu. Technol. **46**(3), 629–648 (2015)
35. Rennie, F., Morrison, T.: E-Learning and Social Networking Handbook: Resources for Higher Education. Routledge, New York (2013)
36. Shelton, C.: Giving up technology and social media: why university lecturers stop using technology in teaching. Technol. Pedagogy Educ. **26**(3), 303–321 (2017)

37. Sobaih, A., Moustafa, M.: Speaking the same language: The value of social networking sites for hospitality and tourism higher education in Egypt. J. Hospitality Tourism Educ. **28**(1), 21–31 (2016)
38. Veletsianos, G., Kimmons, R., French, K.D.: Instructor experiences with a social networking site in a higher education setting: expectations, frustrations, appropriation, and compartmentalization. Educ. Tech. Res. Dev. **61**(2), 255–278 (2013)
39. Wong, B.T.M.: Teaching innovations in Asian higher education: perspectives of educators. Asian Assoc. Open Univ. J. **13**(2), 179–190 (2018)
40. Wong, B.Y.Y., Wong, B.T.M., Pang, S.: A framework for effectiveness of institutional policies on technology-enhanced learning. In: Li, K.C., Wong, T.L., Cheung, S.K.S., Lam, J., Ng, K.K. (eds.) Technology in Education. Transforming Educational Practices with Technology. Communications in Computer and Information Science, vol. 494, pp. 175–186. Springer, Berlin (2015). https://doi.org/10.1007/978-3-662-46158-7_18

The Impact of AWE and Peer Feedback on Chinese EFL Learners' English Writing Performance

Yanna Xie, Liyan Huang(✉), and Yilu Wang

School of Foreign Studies, Language Cognition and Assessment Centre, South China Normal University, No. 55, Zhongshan Avenue West, Guangzhou, Guangdong, China
huangliyan@m.scnu.edu.cn

Abstract. This study compared and contrasted Automated Writing Evaluation system feedback (hereinafter referred to as "AWE feedback") with peer feedback, and obtained the views of the teacher and students of both types. A mixed-method approach was used to assess their impact on the writing performance of 64 second-year non-English-major students in two classes at a college. The scores given to students' assignments were analyzed, with results indicating that both types of feedback can significantly improve students' writing scores. To further explore which aspects of students' writing are impacted, some textual features of the texts were analyzed. We found that AWE feedback significantly improved accuracy, lexical diversity and syntactic complexity, while peer feedback only significantly improved accuracy, but not lexical diversity or syntactic complexity. Classroom observations and interviews were conducted and a questionnaire administered to gain the teacher's and students' perspectives on different types of feedback. The implications for implementing both types of feedback and integrating them into classroom teaching effectively are drawn out.

Keywords: AWE feedback · Peer feedback · Writing performance · Non-English-major student

1 Introduction

Writing is an essential component of students' English proficiency, and feedback has a pivotal role to play in the English as a Foreign Language (hereinafter referred to as EFL) writing classroom. As a central part of teaching Second Language (hereinafter referred to as L2) writing to students, feedback has taken on various forms over the past decades. While a universally agreed taxonomy of feedback strategies in L2 writing is still absent in the literature, some major forms are discernible in terms of their paradigmatic orientations towards writing pedagogy.

Teacher feedback has been acknowledged as one of the most useful types of feedback. However, the use of the process writing method has also led to a trend of replacing teacher feedback with peer feedback, writing workshops, meetings and AWE feedback (Hyland and Hyland 2019).

L.-K. Lee et al. (Eds.): ICTE 2020, CCIS 1302, pp. 258–270, 2020.
https://doi.org/10.1007/978-981-33-4594-2_22

Correspondingly, as an auxiliary means of teacher feedback, research on peer feedback and on AWE feedback in English as Foreign Language (EFL) writing has grown enormously in the past two decades. But most such research investigated either AWE feedback or peer feedback, and some compared either AWE feedback or peer feedback with teacher feedback. There have also been some studies on integrated feedback. The difference between AWE and peer feedback and how the two types can be integrated in the EFL classroom is still an active research field.

In this study we compared and contrast the impact of AWE feedback with that of peer feedback on non-English majors' writing performance, and the participating teacher's and students' views of using such feedback for the revision of their writing. We suggest that both types of feedback can significantly increase the scores given to writing assignments but teacher should make the most of AWE feedback in improving the lexical diversity and syntactic complexity of students' essays.

2 Background

2.1 Progress Writing Theory

In theories of writing and research into writing, the understanding of the essence of writing has undergone a change from "product" to "process". Behaviorist learning theory holds that study is a stimulation-reaction-strengthening process, and so writing is considered a "habit" that requires to be continually developed and enhanced. However, this perspective on writing completely ignores both the objectives and the readers of writing, and thus overlooks the importance of thinking, i.e., of planning. While traditional views of writing overemphasize the product view, good writers focus on thinking about who they are writing for and what they wish to say instead of grammar, which makes writing a combination of process and product. Process writing theory emphasizes the writing process itself and advocates learning to write through writing, with the process divided into three periods, pre-writing, writing, and revision after writing, with emphasis placed on the importance of revision to improve the writing ability of students.

2.2 Peer Feedback

Peer feedback, also called peer review, is an activity used in the L2 writing class (Yu and Hu 2017). Peer feedback has been adopted as a new strategy in teaching writing to students to promote cooperative learning and autonomous learning. In the current study peer feedback refers to the process in which learners give feedback on texts written by other learners in the same learning period, more specifically to the writer of a given text. Peer feedback may include the rating of a text and any comments produced in the process of providing feedback.

A key question with regards to peer feedback research is how effective peer feedback is and what value it offers. Most studies so far have compared the effectiveness of peer feedback with teacher feedback (Yang et al. 2006). Such comparative studies have mainly drawn two conclusions, namely that peer feedback is effective in improving students' learning autonomy (Yang et al. 2006) and in reducing the number of language mistakes

(Diab 2011). Peers are more like "real readers" instead of "judges"; besides, it is thought that peer feedback might provide fresh ideas and suggest expressions different from those given in teacher feedback. It is also thought likely that some of the feedback might be adopted by students to improve their writing quality (Yang et al. 2006).

2.3 AWE Feedback

AWE feedback is "feedback for electronic writing generated by a computer system" (Stevenson and Phakiti 2019, p. 125). The study is concerned with feedback generated by an AWE system instead of feedback provided by teachers or peers, with computers as the medium of writing and receiving feedback. The AWE system has been steadily used in the classrooms of many primary schools, high schools, universities and colleges; and it has been used more and more frequently in the last few years (Stevenson and Phakiti 2014). The biggest and most popular AWE system in China is Pigai.

The key to the study of AWE feedback is being able to verify the reliability and validity of the AWE system regarding consistency between AWE feedback systems and manual feedback systems (Stevenson and Phakiti 2014). The results from a majority of studies show that there is great consistency between AWE feedback systems and manual feedback systems. The AWE feedback system, as a "virtual reader", can provide timely feedback, including student writing statistics, and many opportunities for practice and revision, thus stimulating students' motivation to write and revise; it also spares teachers from a heavy correction workload (Grimes and Warschauer 2010). Further, it is likely to improve students' learning autonomy (Zhang and Hyland 2018). Besides, many studies have noted the positive effect of AWE feedback on students' overall score and error rate.

However, researchers also found that the identification of errors was not as good as in teachers' feedback. While students recognized that the feedback was provided by an AWE system, they showed by their actions that they ascribed more value to teachers' feedback (Dikli and Bleyle 2014). In addition, the AWE system cannot evaluate and promote particular content, ideas, and structures. In fact, in a way it promotes a primarily formalist approach to writing, for example, an essay getting a high score by the AWE system does not mean that it is a good essay. In other words, while the writing of a human being might be complicated and diverse, the feedback of an AWE system is superficial and inflexible. Whether an AWE system can truly evaluate the writing of a human being is up for debate. Besides, some teachers and students think the feedback provided by an AWE system is neither accurate nor clear.

2.4 Integrated Feedback and the Gap

More and more Chinese scholars have focused on the combination of multiple feedback resources to build up an integrated feedback mechanism. The three most popular feedback methods are AWE feedback, peer feedback, and teacher feedback, with four combinations in total: AWE feedback + teacher feedback, peer feedback + teacher feedback, AWE feedback + peer feedback, AWE feedback + peer feedback + teacher feedback. Researchers have been mainly interested in the first two combinations. Although the fourth one has also been investigated (Li 2019), a key question has not been answered in these studies: 1) Why do they pay such attention to the role of AWE feedback in some integrated feedback model, and even claim that their model is based on the AWE system (Li 2019)? Since most studies showed that both AWE feedback and peer feedback tend to focus on surface language errors and neglect content and overall structure, a comparison of the effectiveness of these two feedback methods is needed. Although several studies suggested a combination of peer feedback and teacher feedback based on MicroBlog – a Chinese social media blog (Liu and Wang 2010) – or the peer feedback system Peerceptiv (Gao et al. 2018), these two systems in the two feedback combinations only played a role of an online platform without actual AWE feedback, so a study of these could not be considered a study comparing AWE feedback and peer feedback.

In the study we set out to answer the following three questions:

1) To what extent do AWE and peer feedback influence Chinese EFL learners' English writing scores?
2) To what extent do AWE and peer feedback influence textual features of Chinese EFL learners' English writing?
3) What are teachers' and students' perceptions of AWE feedback and peer feedback?

3 Research Methodology

3.1 Participants and Procedures

The study sample (n = 64) comprised two intact classes of students in the first semester of their second year in a Chinese college, whose native language is Chinese and their teacher—Mr. W. Their ages ranged from 19 to 22 years, and they had not taken part in any kind of peer feedback or AWE feedback before this study. The researchers sought and gained the participants' consent to carry out research on their texts. The original drafts by the students in the two classes were graded by the teacher and one of the researchers (see Sect. 4.1). The average of two reviewers' scores was weighted as the final score of the texts, which was not communicated to the students. To explore whether the overall scores of their original drafts were statistically different, independent-samples *t*-tests were conducted which revealed no statistically significant difference, indicating a similar writing performance at the beginning of the study.

One of the two classes was randomly chosen as the peer feedback group, and the other as the AWE feedback group. As shown in Fig. 1, students were asked to write an application for a campus club or society before class started to minimize the time pressure for students (Yu and Lee 2014). The test material was chosen by the teacher according to her teaching schedule to keep the class as close to the natural writing class as possible. After some simple training, the students received feedback from either the AWE system or their peers, after which they revised their draft according to the feedback. In the training of the peer feedback group students could choose a pair partner they trusted; they were also allowed to use their L1 so they could "more fully participate in developing their ideas for writing" (Pennington and Yue 1996, p. 243). In the half-hour training, the teacher modelled giving feedback on a student's draft (Chang 2015), with the help of the peer feedback sheet and using the think-aloud method, then students reviewed a sample paragraph for practice following the teacher's instruction. After training, each student was asked to review two drafts in the space of 20 min. In the first 10 min training of the

AWE feedback group, the teacher instructed students to download the Pigai application and log in; then students were given 15 min to enter their texts into the system. The teacher then demonstrated how to revise their draft according to the feedback, which took about five minutes.

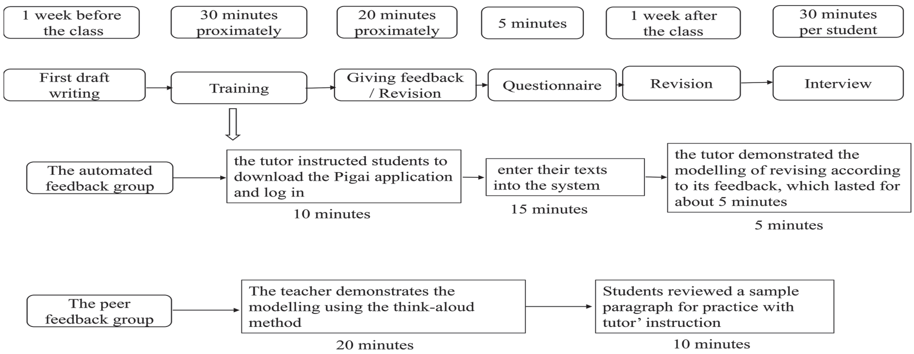


Fig. 1. Procedure of the quasi-experiment and training

Students in both groups were told that in the revision stage they could decide whether to follow the advice received as part of the feedback or not and that they could disregard it if it did not fit their texts. Students were allowed one week to revise their original draft. Survey questionnaires were distributed, filled out, and collected in class. Semi-structured interviews with the teacher and four students (two higher-level and two lower-level students) in each group were conducted after they had submitted their final drafts. Each interview lasted approximately half an hour and was conducted in Chinese.

3.2 Instruments

As mentioned above, the current study adopted a combination of quantitative and qualitative data collection methods. Data, including students' original drafts and revised drafts, teachers' scores for these texts, and textual feature indices of these texts, were collected for quantitative analysis, and classroom observations and interviews with the teacher and students were conducted for qualitative analysis.

AWE System. Pigai (literally meaning "correction" in Chinese) is an AWE system application designed for assessing the quality of writing by Chinese learners of English. It generates a holistic score, general feedback, sentence-based analysis and class ranking, all in Chinese (see Fig. 2 for a snapshot of the English version).

AWE scores are based on the extent to which a student's text aligns with a large corpus of human-scored essays on 192 dimensions (e.g., average sentence length and the number of conjunctions). The system provides 11 holistic score scales for task-setters to choose from, ranging from 0–5 to 0–100 points, but no scoring rubrics are available online. General feedback covers several dimensions such as vocabulary, sentence, structure and organization, and content relevance, with a bar graph illustrating the relative strength or weakness of the text in each of these areas. The sentence-based analysis includes

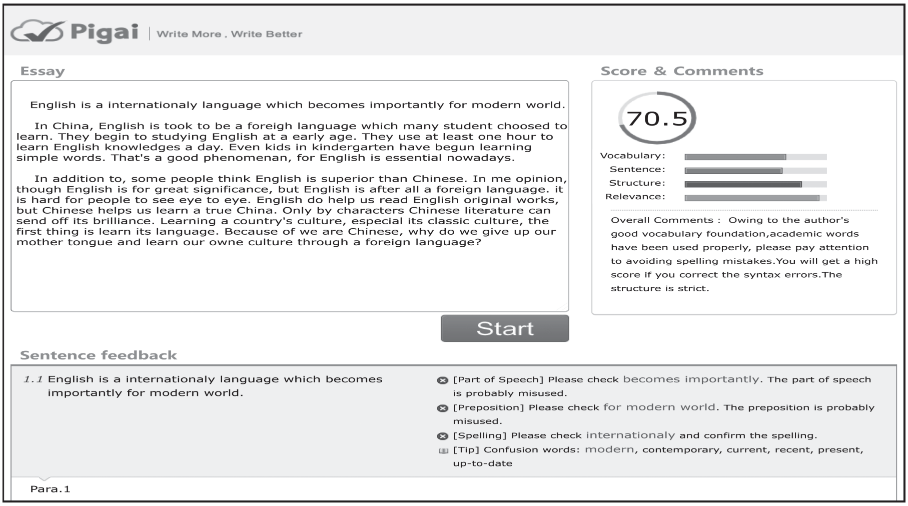


Fig. 2. Snapshot of feedback from Pigai

feedback on collocations (based on a corpus), word choices, verb forms, tenses, spelling, punctuations, etc. for every single sentence (see Fig. 2). The sentence-based analysis comprises mostly corrective feedback accompanied by further information regarding the dimensions it checks.

Measures of Textual Features. Foster and Skehan (1996) proposed a now widely accepted model for measuring the textual features of essays covering three aspects: accuracy, fluency and complexity, with the complexity able to be further divided into lexical diversity and syntactic complexity, a model adopted by many studies. As numerous studies show the importance of cohesion (McNamara et al. 2014), cohesion is analyzed as well.

The researchers entered 64 original and revised drafts of the peer feedback group one by one into Coh-Metrix 3.0 to calculate lexical diversity, syntactic complexity and fluency, and then analyzed the 64 original and revised drafts of the AWE feedback group in the same way. The Coh-Metrix tool is a web-based AWE textual assessment tool which can calculate multiple characteristics of a text in seconds. Its public version provides 11 measures and 106 indices, of which four measures and six indices were chosen for the present study, along with one measure and one index rated by a human:

(1) Complexity

Complexity is the writer's ability to use varied complex structures and vocabulary. It is generally believed that it consists of syntactic complexity and lexical diversity.

(a) lexical diversity

Lexical diversity is the variation and sophistication of the words in a text (Lu 2012). We adopted the type:token ratio (LDTTRc, LDTTRa) incorporated in Coh-Metrix. A low type:token ratio indicates that words are repeated many times, which means lower lexical diversity (McNamara et al. 2014).

(b) Syntactic complexity

Syntactic complexity refers to the degree of sophistication and variation of the structures produced. We adopted two indices from Coh-Metrix: sentence similarity of all combinations across paragraphs (SYSTRUT) and left embeddedness (SYNLE). Generally, a lower SYSTRUT or more SYNLE is associated with higher syntactic complexity (McNamara et al. 2014).

(2) Fluency

Fluency refers to the writer's ability to write fluently. A measurement of fluency has not been generally agreed but many researchers use Length (DESWC) to measure fluency. In writing, greater length generally means greater fluency.

(3) Cohesion

Latent Semantic Analysis (LSA) is a useful way to measure a text's cohesion. It provides measures of semantic overlap between sentences or between paragraphs. LSA similarity between all possible pairs of sentences in a paragraph (LSASSp) and LSA Given-New (LSAGN) estimate the proportion of new information in each sentence. LSA indices range from 0 to 1, with a higher value indicating greater cohesion (McNamara et al. 2014).

(4) Accuracy

Accuracy was mainly measured by the ratio of the number of errors to the number of all words (E/W) (Chen 2010). An increase in E/W generally indicates lower accuracy. After two researchers had counted the number of errors in each text, the number of all words of each text was calculated using Word 2003, and the E/W was calculated using Excel 2003.

SPSS 22.0 was used to statistically analyze the data collected.

4 Results and Discussion

4.1 Results of Scores of Two Groups

The improvement of students' revised drafts over the original drafts was measured based on the teacher's grades. The teacher and one of the researchers graded the students' writing, using the same scoring guide but without being told which was the original draft and which the final one.

Students' writing scores fit the normal distribution and their homogeneity of variance ($p > 0.05$) was observed. To primarily see the influence of peer feedback and AWE feedback on students' writing performance, paired-sample *t*-tests were used separately (see Table 1) to compare the mean scores of the original and revised drafts for both the peer feedback group and the AWE feedback group. The difference was statistically significant for both groups ($t_{\text{peer}} = 0.016$, $p_{\text{peer}} < 0.05$, $t_{\text{AWE}} = 0.000$, $p_{\text{AWE}} < 0.05$), which indicates that both peer feedback and AWE feedback helped to improve the overall score of students' writing. However, the mean score of the AWE feedback group increased by 2.8516, remarkably more than that of the peer feedback group with 0.8723. The results show that AWE feedback works better in improving students' overall writing scores.

Table 1. Paired sample *t*-tests of scores of two groups.

	Scores of original drafts		Scores of revised drafts		MD	t	df	Sig. (2 tailed)
	M	SD	M	SD				
Peer feedback (N = 32)	81.7813	8.24419	82.6536	7.69943	1.95101	−2.537	31	0.016
AWE feedback (N = 32)	81.125	7.59669	83.9766	7.30785	1.83477	−8.793	31	0.000

4.2 Results of Textual Features of Two Groups

To further see in which aspects do peer feedback and AWE feedback influence students' writing performance, textual features of students' texts were analyzed and compared. All texts were analyzed by Coh-Metrix tool 3.0 and Excel 2003 for measures of textual features. Then the data were processed using SPSS 22.0 to see in which aspects two types of feedback helped to improve students' writing. For all indices chosen, homogeneity of variance (*p* values range from 0.060 to 0.890) was observed. Table 2 displays the results of paired sample t-tests of textual features of two groups.

LDTTRc = Type:token ratio of content words; LDTTRa = Type:token ratio of all words; SYSTRUT = Sentence similarity of all combinations across paragraphs; SYNLE = Left embeddedness; LSASSp = LSA similarity between all possible pairs of sentences in a paragraph; LSAGN = LSA Given-New; DESWC = Length; E/W = the ratio of the number of errors to the number of all words.

In terms of lexical diversity, LDTTRc and LDTTRa of the original and revised drafts both showed a statistically significant increase ($t_{\text{LDTTRc}} = -1.913$, $p_{\text{LDTTRc}} = 0.048$, $t_{\text{LDTTRa}} = -2.332$, $p_{\text{LDTTRa}} = 0.025$) in the AWE feedback group. LDTTRc and LDTTRa of the original and revised drafts in the peer feedback group showed a decrease in mean but this was not statistically significant. These results, therefore, show that AWE feedback can significantly improve students' lexical diversity in a short time but peer feedback cannot.

The paired-sample *t*-test of SYNSTRUT ($t = 2.074$, $p = 0.046$) revealed a statistically significant decrease between the original and the revised draft in the AWE

Table 2. Paired sample *t*-tests of textual features of two groups.

Measures	Indices	Peer feedback group (N = 32)			AWE feedback group (N = 32)		
		MD	t(N = 31)	Sig. (2 tailed)	MD	t(N = 31)	Sig. (2 tailed)
Lexical diversity	LDTTRc	−0.0075	0.988	0.371	0.0081	−1.913	0.048
	LDTTRa	−0.0066	1.079	0.311	0.0106	−2.332	0.025
Syntactic complexity	SYSTRUT	−0.0969	0.38	0.221	−0.1481	1.613	0.046
	SYNLE	−0.0006	0.089	0.283	−0.0059	1.617	0.231
Cohesion	LSASSp	0.0077	−0.559	0.58	0.0115	−0.515	0.61
	LSAGN	0.0056	−1.194	0.242	0.0000	0.004	0.997
Fluency	DESWC	1.5313	−0.425	0.674	2.4375	−0.764	0.439
Accuracy	E/W	−0.0112	4.334	0.000	−0.0129	4.378	0.000

group. The mean of SYNLE increased as well, although this was not statistically significant. These results indicate that students' syntactic complexity increased significantly in the AWE feedback group, while no statistically significant difference was found in the peer feedback group. So compared to peer feedback, AWE feedback worked better in facilitating students' use of more complex syntactic structures.

Our findings regarding lexical diversity and syntactic complexity are in line with Ruegg's (2015) and Kellogg et al.'s (2010) studies. Ruegg (2015) compared the effectiveness of teacher feedback with peer feedback. In his study, the lexical diversity and syntactic complexity were covered as subcategories at the levels of lexis and grammar respectively, without significant improvement at either level in the peer feedback group. Kellogg et al. (2010) found that continuous AWE feedback helped decrease word repetition, which is equal to improving lexical diversity in the current study.

No statistically significant difference was found for LSASSp or LSAGN, either in the peer feedback group or the AWE feedback group, and the increase in means was minor. Since LSA and cohesion are at a higher level of a putative model of language than words and sentences and are not easy to improve or measure, we can only say that these two types of feedback cannot increase students' cohesion in a short period of time.

In terms of fluency, the AWE feedback group showed a larger increase than the peer feedback group, but neither group showed a statistically significant increase. This may be due to the number of words required by the teacher. Kellogg et al. (2010) found only a marginally significant improvement in the group given intermittent AWE feedback, but no improvement in the group given continuous AWE feedback. This is strange, but the authors did not give a clear explanation of this result.

The *t* statistics and the *p* values in Table 2 reveal that the error rates of the two groups decreased significantly ($t_{\text{peer}} = 4.334$, $p_{\text{peer}} = 0.000$, $t_{\text{AWE}} = 4.378$, $p_{\text{AWE}} = 0.000$), indicating that both AWE feedback and peer feedback helped students to reduce errors. This is in line with previous studies (Diab 2011; Kellogg et al. 2010).

In summary, AWE feedback and peer feedback can significantly increase the scores and the accuracy of students' essays, but AWE feedback was more successful in helping students improve the lexical diversity and syntactic complexity of their essays.

4.3 Results of Classroom Observations, Questionnaires and Interviews

When being asked to what extent they thought the AWE feedback or peer feedback helped to improve the quality of their writing, 86.6% of students in the AWE feedback group chose "useful" or "very useful", but only 46.7% of students in the peer feedback group made the same choice. In the interview, most students in the AWE feedback group considered the AWE system a reliable reviewer in detecting their linguistic errors and said that they usually take its advice if they can understand it. Furthermore, many students mentioned that the sentence-based analysis provided was very specific and instructive. However, the teacher thought that there were misidentifications of errors by the AWE feedback. For example, it always identified "Mr." or "Ms." as a punctuation error. She explained that to prevent students from being misled she tested the system in advance to find bugs and gave explicit instructions in class, reminding students to ignore those suggestions. Students in the peer feedback group were of the view that much feedback given by their peers did not fit their texts. A student in the peer feedback group did not make any changes in the revision stage. His original draft got 89 points. A higher-level student emphasized that it was incorrect feedback which made her lose faith in peer feedback. However, the teacher believes that peer feedback will help to promote the development of students' critical thinking in the long term.

Most students had a positive attitude towards AWE feedback, but some feedback given by the AWE system was too advanced for some students, and the teacher also expressed concern about its reliability. Classroom observation found that students in the AWE feedback group asked the teacher more than fifteen times for explanations of terms in feedback. One student said that he could not understand the comment "The expression 'will very hard work' is wrong, the modal verb should be followed by the root form of verbs." He corrected his error after the teacher explained what modal verb and the root form of the verb referred to in this expression. Similarly, a lower-level student mentioned in his interview that "there are many terms I don't understand, like 'definite article', and I just ignore most of the additional information it offers, because I don't understand it and it's a waste of time." These issues should be brought to the attention of the developers of Pigai to be rectified since an increase in accuracy would be the biggest attraction for students. Especially lower-level students would benefit because they are more likely to make linguistic mistakes.

Students in the peer feedback group understand most of the feedback without difficulty, but some have trouble giving feedback, and higher-level students seem to improve less from lower-level students' feedback. In the peer feedback group, there was no demand for an explanation of the feedback received, and students' questions were more about what to do and how to give comments to their peers. A boy came to the teacher for help with translating one of his comments into English. In an interview, a student remarked that "It's new for me to give feedback, but to be honest, it's not easy." His dilemma was echoed by another student's statement: "although the teacher had provided a model of how to give feedback, it was hard to get started, so I hesitated for a long time."

The classroom observations, questionnaires and interviews showed that students had a more positive attitude towards AWE feedback than towards peer feedback, and that AWE feedback seemed to be more effective in improving the learning autonomy of students, a finding in line with Zhang and Hyland's (2018) study. However, the written feedback given by the AWE system in students' L1 is difficult for lower-level students, so the guidance provided by teachers is very important. Students, especially higher-level students, have some concerns about the reliability of peer feedback, and in turn giving useful feedback demands a higher level of language proficiency from students that some simply may not have.

There are several factors that should be taken into consideration by teachers when designing the teaching of writing, instead of following a fixed formula of integrated feedback. For example, in which aspects of writing does the teacher want students to improve? Further, teachers should ask themselves whether students are likely to be able to understand and utilize the feedback the system or peers provide. Lastly, what are the characteristics of different types of feedback that are of most use to students? And while teachers should be able to help students build confidence in the chosen type of feedback (Wang 2014; Zhao 2010), the developers of AWE systems should ensure that their systems provide feedback in language that is targeted for students at different levels.

5 Conclusion

In this study, we have sought to compare and contrast the impact of AWE feedback with that of peer feedback on non-English majors' writing performance, and the participating teacher's and students' views of using such feedback for the revision of their writing. Our findings showed that both AWE feedback and peer feedback can significantly increase the scores given to writing assignments but they work differently in influencing the students' choice of the grammatical and lexical features of their essays. Students had a more positive attitude towards AWE feedback than peer feedback. Further, AWE feedback was found to be more effective in improving the learning autonomy of students (Zhang and Hyland 2018). However, the written feedback given by the AWE system via Pigai is difficult for lower-level students to understand, which makes the guidance provided by the writing teacher all the more important (Liao 2016). On the other hand, having peers provide feedback requires a higher level of language proficiency of them than some students have.

The paper contributes to the literature by providing an empirical study that specifically compares the effects of AWE feedback with those of peer feedback in the teaching of L2 writing. We feel confident that our findings will be able to inform L2 writing teachers' decisions when selecting and designing feedback. We suggest writing teachers choose proper types of feedback according to their teaching objectives. For example, if the emphasis of the class is in improving the lexical diversity or syntactic complexity of students' writing, AWE feedback may be a better option, but training in understanding and using AWE feedback should be provided. Meanwhile, AWE feedback systems should develop more user-friendly methods of reporting misidentifications of specific tasks. Since some teachers may ascribe more value to peer feedback in improving overall scores of students' writing, if peer feedback is chosen, measures should be taken to foster a positive attitude towards peer feedback in students, and, students should be trained to give better and more useful feedback to their peers.

Certain limitations in this study must be acknowledged. Firstly, students' individual differences were not considered in this study. How students' language proficiency and their readiness may affect the way they respond to feedback received warrants further study. Secondly, all AWE feedback in the present study was generated by a single system, examination of other AWE systems will contribute to a more comprehensive understanding of the effectiveness of AWE feedback.

Acknowledgement. This work was supported by a grant from the National Social Science Fund of China (15BYY080) to the corresponding author Liyan Huang.

References

Chang, C.Y.H.: Teacher modeling on EFL reviewers' audience-aware feedback and affectivity in L2 peer review. Assessing Writ. **25**, 2–21 (2015)

Chen, H.Y.: A study of the writing performance measures and their attributes. Mod. Foreign Lang. **33**(1), 72–80 (2010)

Diab, N.M.: Assessing the relationship between different types of student feedback and the quality of revised writing. Assessing Writ. **16**(4), 274–292 (2011)

Dikli, S., Bleyle, S.: Automated essay scoring feedback for second language writers: how does it compare to instructor feedback? Assessing Writ. **22**, 1–17 (2014)

Foster, P., Skehan, P.: The influence of planning and task type on second language performance. Stud. Second Lang. Acquisition **18**, 299–323 (1996)

Gao, Y., Zhang, F.H., Zhang, S.J., Christian D.S.: Effects of receiving peer feedback in english writing: a study based on Peerceptiv. In: Media in Foreign Language Instruction, vol. 67, pp. 3–9 (2018). 000(002)

Grimes, D., Warschauer, M.: Utility in a fallible tool: a multi-site case study of automated writing evaluation. J. Technol. Learn. Assess. **8**(6), 44 (2010)

Hyland, K., Hyland, F. (eds.): Feedback in Second Language Writing: Contexts and Issues. Cambridge University Press, Cambridge (2019)

Kellogg, R.T., Whiteford, A.P., Quinlan, T.: Does automated feedback help students learn to write? J. Educ. Comput. Res. **42**(2), 173–196 (2010)

Li, G.F.: Study on the influence of AWE-based integrated feedback on English composition revision. Foreign Language Education, 0(2), 67 (2019)

Liao, H.-C.: Using automated writing evaluation to reduce grammar errors in writing. ELT Journal **70**(3), 308–319 (2016)

Liu, Y., Wang, X.L.: Multifeedback in blog-based college English writing. Foreign Language Education **031**(004), 70–72 (2010)

Lu, X.: The relationship of lexical richness to the quality of ESL learners' oral narratives. Mod. Lang. J. **96**(2), 190–208 (2012). https://doi.org/10.1111/j.1540-4781.2011.01232_1

McNamara, D., Graesser, A., McCarthy, P.M., Cai, Z.: Automated evaluation of text and discourse with Coh-Metrix. Cambridge University Press, Cambridge (2014). https://doi.org/10.1017/CBO9780511894664

Pennington, M.C., Yue, F.: Explaining Hong Kong students' response to process writing: an explanation of causes and outcomes. J. Second Lang. Writ. **5**(3), 227–252 (1996)

Ruegg, R.: The relative effects of peer and teacher feedback on improvement in EFL students' writing ability. Linguistics and Education **29**, 73–82 (2015)

Stevenson, M., Phakiti, A.: The effects of computer-generated feedback on the quality of writing. Assessing Writing **19**, 51–65 (2014)

Stevenson, M., Phakiti, A.: Automated feedback and second language writing. In: Feedback in Second Language Writing: Contexts and Issues, pp. 125–142 (2019)

Wang, W.: Students' perceptions of rubric-referenced peer feedback on EFL writing: a longitudinal inquiry. Assessing Writ. **19**, 80–96 (2014)

Yang, M., Badger, R., Yu, Z.: A comparative study of peer and teacher feedback in a Chinese EFL writing class. J. Second Lang. Writ. **15**(3), 179–200 (2006)

Yu, S., Hu, G.: Understanding university students' peer feedback practices in EFL writing: insights from a case study. Assessing Writing **33**, 25–35 (2017)

Yu, S., Lee, I.: An analysis of Chinese EFL students' use of first and second language in peer feedback of L2 writing. System **47**, 28–38 (2014)

Zhang, Z., Hyland, K.: Student engagement with teacher and automated feedback on L2 writing. Assessing Writing **36**, 90–102 (2018)

Zhao, H.: Investigating learners' use and understanding of peer and teacher feedback on writing: a comparative study in a Chinese English writing classroom. Assessing Writing **15**(1), 3–17 (2010)

A Review of Small Private Online Courses in Higher Vocational Colleges

Ning-wei Sun[1] and Simon K. S. Cheung[2](✉)

[1] Guangdong Women's Polytechnic College, Guangzhou, China
sun1990917@163.com
[2] The Open University of Hong Kong, Hong Kong, China
kscheung@ouhk.edu.hk

Abstract. Small private online courses or SPOC are localized instances of massive open online courses that are intentionally designed for on-campus learning purposes. Through a review of three SPOC in a higher vocational college in Guangzhou, this paper studies the effectiveness of SPOC in terms of learning participation and teacher-student interaction. The current practices are analyzed. It is revealed that, while the teachers actively promote the learning interaction through the SPOC, the student initiatives to share knowledge may be ignored. Therefore, the teachers should devise a teaching strategy that can help stimulate the students' learning motivation, balancing their learning needs and interests. It is also revealed that learning participation and teacher-student interaction would be improved if the students' learning differences can be timely catered. Therefore, homogeneous grouping of students by their learning needs and abilities are recommended. These provide a useful reference for the implementation of SPOC in higher vocational colleges.

Keywords: SPOC · MOOC · Learning participation · Teaching-student interaction · Teaching strategy · Vocational education

1 Introduction

Open online courses are one major type of open educational resources widely used in educational institutions at different levels, especially higher education [1–3]. Massive open online courses (MOOC) are representative examples of open online courses [4, 5]. The effectiveness of open online courses, or open educational resources in general, have been well perceived by both full-time and distance-learning university students [6–11]. In recent years, Small Private Online Courses (SPOC) has evolved from Massive Open Online Courses (MOOC) as a localized instance, intentionally for on-campus teaching and learning purposes [12, 13]. Different from MOOC, SPOC focus on certain and smaller groups of students who are ready for interact with each other. Interaction is essential to online education, and SPOC enhances learning participation and interaction [14, 15]. SPOC can be used in a blended learning environment as some small online courses to supplement on-campus classroom teaching and learning, and the effectiveness is promising [16, 17].

L.-K. Lee et al. (Eds.): ICTE 2020, CCIS 1302, pp. 271–280, 2020.
https://doi.org/10.1007/978-981-33-4594-2_23

During the implementation of SPOC, MOOC resources are selected for use in a teaching and learning environment combining flipped classroom and project-based learning, where students are provided with high-quality online and offline learning resources. In recent years, MOOC have been expanding in a compound rate. From a vast amount of MOOC, many universities and colleges in China have started to select those high-quality open online courses for SPOC. These college-level online open course courses are further selected for use at the provincial level and national level, as shown in Fig. 1.

Fig. 1. Online course construction process

At present, a large number of high-quality online open courses have been selected for use as SPOC in China. Researchers and practitioners are actively investigating the effective practices of teaching and learning with SPOC that can improve the students' learning participation and interaction. This paper analyzes the current practices in China through a review of three SPOC in a higher vocational college in Guangzhou. The focus is placed on learning participation and interaction. Based on the findings, some strategies are suggested.

The rest of this paper is structured as follows. Section 2 states the research question. Section 3 describes the research design and implementation, where three SPOC are reviewed and the current practices are analyzed. Section 4 concludes this paper with discussion on the analysis results and findings. Some effective teaching and learning practices with SPOC are recommended.

2 Motivation of This Study

According to Ai et al., the core factors affecting the implementation effect of SPOC includes the learning resources, learning environment and teacher-learner interaction [18]. Yang & Yu analyzed the effects of SPOC teaching from the teaching videos, learning packs, and learning participation and interaction [19]. Liao & Liu also investigated the core factors affecting the SPOC learning effect in colleges and universities from the teaching management, learning materials, learning participation and learning environment [20]. The course development should better be driven by pedagogies and student needs [21, 22].

Summarizing the findings from the literature, there are external factors and internal factors, affecting the effectiveness of SPOC implementation. External factors mainly include the teaching and learning contents (i.e. teaching videos, learning packs and other learning resources), teaching management, and learning environment. Internal factors include teachers, students, learning participation and interaction. In the current implementation of SPOC in higher vocational colleges in China, the external factors are relatively unchanged. For a higher vocational college, the same online learning platform is used, while the learning contents are developed from online open courses. There is often no change on the teaching management system within the college. On the other hand, the internal factors such as learning participation and teacher-student interactors are essential to the SPOC implementation.

At present, relevant research of SPOC mainly focuses on the teaching methods and practices, and the factors influencing the effectiveness. There is a lack of studies on analyzing learning participation and teacher-student interaction. In recent years, many higher vocational colleges in China have been facing a challenge that the learning motivation is insufficient and the learning participation is low. This motivates the authors to investigate how SPOC can be effectively implemented to enhance learning participation and promote teacher-student interaction.

3 Research Design and Implementation

In this study, the learning interaction data of three SPOC in Guangdong Women's Polytechnic College, Guangzhou, China are collected. These courses are:

- *Digital Camera Technology*,
- *3D Animation Design,* and
- *Practical Japanese Translation.*

The learning interaction data would be analyzed, and the current teaching practices in the college would be reviewed.

3.1 Data Source

Components of the three selected SPOC courses are collected as the sample for our study. These three courses exhibit a number of characteristics which are highlighted as follows.

First, the course contents and learning materials are relatively stable, after several rounds of updating. The course design and teaching methods are quite established. Second, according to the student evaluation, the teachers of these courses have been ranked within the top 15% of all teachers for three consecutive semesters. In other words, the students have given favourable evaluation on the courses (course contents, course design, teaching methods, etc.) and the teachers for three consecutive semesters. These also represent that the learning satisfaction of these course have been high. Third, not only carrying out as SPOC in the colleges, the courses have also been used as a MOOC on a social network platform for quite a while. Fourth, the courses have been rated as ones among the excellent online open courses in the Guangdong Province.

3.2 Data Collection and Organization

Interactive Data Collection and Organization of Learners in Online Learning. We take the online learning interaction data of one semester in three SPOC courses from 2018 to 2019 for the study. These data mainly cover the collaborative discussion between students in the project-based learning component, the interaction between teachers and students in the homework component, and the interaction both between teachers and students and between students in the discussion component.

Making reference to relevant knowledge construction models in the literature [23–26], this paper use the 5-level interaction model proposed by Xiong [27]. Table 1 shows the interaction model, showing each interaction level and its definition. The learning interaction data collected from the three SPOC courses would be selected and classified according to this model.

Table 1. Interaction five-level interaction model of online learning

Interaction level	Definition and meaning
Information sharing	Students share information, describe the problems and put forward ideas (published smart products openly) for the discussed themes in a specific problem context
Deepened understanding	Students find inconsistencies with the personal construction of knowledge through analyzing various views, and further obtain conclusions through the questioning and debate, thus indicating the position or raising objections to deepen the understanding of the problems
Meaning negotiation	Students carry out the collaborative construction of group knowledge through the meaningful social consultation and significance clarification
Inspection and modification	Students inspect and revise the newly constructed collaborative points, making the individual acquire relevant knowledge during this process and internalize into the cognitive schema, in order to better assimilate or adapt the individual's construction knowledge
Creation and application	Students create and apply new knowledge based on the reached consensus

Interview Data Collection and Organization of Teachers. Moore divided the learning interactions into three categories based on different interaction objects, namely, the interaction between students and course contents, interaction between students and teachers, and interaction between students [28]. Following this categorization, the interview should focus on the course contents, teacher-student interaction, and interaction between students. Interviews are carried and analyzed using the model by Li & Xie [29].

3.3 Data Analysis

Interaction Data Analysis of the SPOC Learner. From the collected interaction data sample, a total of 728 valid interaction items are identified from the three SPOC courses are categorized. They are classified according to the 5-level interaction model. Table 2 shows the number of interaction items for each of the SPOC courses and all three SPOC courses by the interaction level. It is shown that the learning interaction tends to realize inspection, modification, creation and application, but there is a lack of interactions on information sharing and deepened understanding.

Table 2. Interaction level data among students in three courses

Interaction levels in online learning	No. of interaction items among students			
	3D animation design	Digital camera technology	Practical Japanese translation	Total
Information sharing	18	22	36	76
Deepened understanding	17	12	31	60
Meaning negotiation	41	30	52	123
Inspection and modification	92	87	121	300
Creation and application	77	62	30	169

Analysis of Teachers' Interview Data. The paper divides the interview contents into three phrases. The first phase or first-level coding is the open coding process, where the original interviews with the teachers are transcribed. The final transcript contains 161 sentences. Each sentence is given a concept, and similar concepts are categorized. After this coding process, the 161 concepts are categorized under 11 categories, namely, maintaining the learning motivation, informing curriculum system, creating display abilities opportunities, hierarchical teaching, combination between the work and learning, project cooperation, real-time discussion on difficult problems, students' participation in course content generation, offline core problem solving, teaching model reform, and the adjustment of online and offline teaching relationships.

Table 3 illustrates this open coding process, where a sentence of the transcript is given a concept which is then categorized.

In the second phase or second level coding (axial coding), the 11 concept categories are further categorized into four major categories, namely, promoting participation in the course learning (maintaining the learning motivation), promoting the interaction between students (creating the opportunities of demonstrating the learning abilities and project cooperation), promoting the effective interaction between teachers and students

Table 3. Open coding process

Sentence example	Concept	Category
During the course implementation process, students like cases related to the actual positions, so I have specially designed ten cases with the gradual increased difficulties to drive them to actively participate in learning	Actual case driven	Maintaining learning motivation
As students become very interested in plots-oriented people design in the 3D animation design, I will provide some story plots they are interested in for them to design their favorite characters	Interest-driven learning	Maintaining learning motivation
The Japanese translation materials I choose for each unit can always connect with the hot topics that those students have recently paid attention to	Continuous attracted attention	Maintaining learning motivation
I would design the related learning project tasks according to the actual requirements and tasks of the animation company for 3D animation designers	Work project design	Combination of working and study
I sometimes provide students with some part-time translation tasks. In this process, I am the gatekeeper to be responsible for the tutoring and optimization, and finally give students some rewards	Part-time translation task	Combination of working and study
I will take students to the TV station to participate in the recording of real programs	Participating in the practical work	Combination of working and study
……	……	

(including the real-time discussion on difficult problems, and offline core problem solving), SPOC practice innovation (students' participation in developing the course contents, making adjustment on online and offline teaching relationships, and improving the teaching practices).

In the third phase or third level coding (core coding), it can be found that the four major categories can be generalized as the "learning interaction and practical innovation in SPOC". Therefore, it is devised that the successful implementation should help improve the students' participation, and promote the positive interaction between teachers and students, and interaction between students.

Table 4 summarizes the first-level coding (open coding), second-level coding and third-level coding process, illustrating how the 161 sentences from the interview transcript are analyzed, and how the concepts are categorized.

Table 4. Three-level coding process

Sentences	Concepts	1st level coding (open coding)	2nd level coding (axial coding)	3rd level coding (core coding)
161 sentences, such as "the Japanese translation materials I choose for each unit always connect with the hot topics that those students have recently paid attention to"	161 concepts, such as continuous attracted attention	Maintained learning motivation	Promoting the participation of course learning	Learning interaction and practical innovation in SPOC
		Informed curriculum system		
		Combination of working and study		
		Creation of display abilities	Promoting the interaction between students	
		Project cooperation		
		Real-time discussion on difficult problems	Promoting the effective teacher-student interaction	
		Solving core problems offline		
		Hierarchical teaching	SPOC practice innovation	
		Adjusting online and offline teaching relationship		
		Students' participation in course content generation		
		Teaching model reform		

Coding Reliability Test. It is necessary to ensure the authenticity and reliability throughout the coding process. During the coding process, different events and concepts are compared. After the coding process, the interviewed teachers were invited to confirm the validity of the coding of their interview answers. They confirmed that the coding accurately denotes their interview answers.

4 Discussion and Conclusion

Based on the review of the 3 SPOC courses and the analysis of the interaction data and interview results, some effective teaching and learning practices with SPOC are recommended and discussed as follows.

4.1 Promoting the Interaction Between Students

It is found from the interaction data that teachers actively promote the learning interaction with the SPOC learning resources, while ignoring the students' initiative to share information and deepen mutual understanding of the course contents. From the interviews with teachers, the most commonly used way for teachers to promote interactions between students is to adopt project-based learning, especially through some related vocational skill competitions and show-case activities inside and outside the campus. More interactions between students for the teamwork or project work are encouraged.

4.2 Promoting the Learning Participation in Multiple Ways

According to Fogg, behavior is usually generated due to three factors, specifically including the motivation, ability and trigger [30]. In order to better promote students' learning participation, it is essential to stimulate their learning motivation, combining with the learning needs and interests. These should be considered in designing the overall curriculum design, defining the course objectives, and setting clear nodes triggering learning activities. In the learning process, students can better understand the learning purposes, learning needs, course objectives, and learning activities and requirements.

4.3 Promote the Effective Interaction Between Teachers and Students

Owing to the separation of SPOC online and offline teaching, teachers are required to timely discuss some difficult problems with students, and help solve the common problems raised by the students in their interaction and discussion. This can remove barriers on some core issues during the learning process.

4.4 Deepen the Interaction Level Through Innovative Practice

Teaching and learning activities can be designed to cater for the learning difference among students. For SPOC teaching, students can be grouped in accordance with their learning abilities and learning needs. This homogeneous grouping of students can help teachers focus on the specific learning needs and learning difficulties of specific groups of students. For example, more difficult or creative tasks can be assigned to those student groups with better learning abilities. Besides catering for learning needs, this homogeneous grouping of students also promotes effective interaction between teachers and students.

References

1. Commonwealth of Learning: Guidelines for Open Educational Resources in Higher Education. UNESCO and Commonwealth of Learning (2015)
2. Cheung, S.K.S., Li, K.C., Yuen, K.S.: An overview of open education resources for higher education. In: Lam, J., Li, K.C., Cheung, S.K.S., Wang, F.L. (eds.) ICT 2013. CCIS, vol. 407, pp. 26–34. Springer, Heidelberg (2013). https://doi.org/10.1007/978-3-642-45272-7_3

3. Cheung, Simon K.S.: A study on the university students' use of open educational resources for learning purposes. In: Cheung, Simon K.S., Jiao, J., Lee, L.-K., Zhang, X., Li, K.C., Zhan, Z. (eds.) ICTE 2019. CCIS, vol. 1048, pp. 146–155. Springer, Singapore (2019). https://doi.org/10.1007/978-981-13-9895-7_13
4. Gable, M.: MOOCs: massive open online courses. EUA Occasional Papers, European University Association (2013)
5. Waks, L.J.: The Evolution and Evaluation of Massive Open Online Courses: MOOCs in Motion. Palgrave Macmillan, New York (2016)
6. Hew, K.F., Cheung, W.S.: Students' and instructors' use of massive open online courses (MOOCs): motivations and challenges. Educ. Res. Rev. **12**, 45–58 (2014)
7. Liu, M., Kang, J., McKelroy, E.: Examining learners' perspective of taking a MOOC: reasons, excitement, and perception of usefulness. Educ. Media Int. **52**(2), 129–146 (2015)
8. Zhang, M., Yin, S., Luo, M., Yan, W.: Learner control, user characteristics, platform difference, and their role in adoption intention for MOOC learning in China. Australas. J. Educ. Technol. **33**(1), 114–133 (2017)
9. Cheung, S.K.S.: Distance-learning students' perception on the usefulness of open educational resources. In: Cheung, S.K.S., Kwok, L.-f., Ma, W.W.K., Lee, L.-K., Yang, H. (eds.) ICBL 2017. LNCS, vol. 10309, pp. 389–399. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-59360-9_34
10. Cheung, S.K.S.: Perceived usefulness of open educational resources between full-time and distance-learning students. In: Cheung, S.K.S., Kwok, L.-f., Kubota, K., Lee, L.-K., Tokito, J. (eds.) ICBL 2018. LNCS, vol. 10949, pp. 357–367. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-94505-7_29
11. Tsang, E.Y.M., Yuen, K.S., Li, K.C., Cheung, S.K.S.: Designing Open Textbooks for Effective Teaching and Learning. In: Lam, J., Li, K.C., Cheung, S.K.S., Wang, F.L. (eds.) ICT 2013. CCIS, vol. 407, pp. 43–55. Springer, Heidelberg (2013). https://doi.org/10.1007/978-3-642-45272-7_5
12. Fox, A.: From MOOCs to SPOCs. Commun. ACM **56**(12), 38–40 (2013)
13. Alario-Hoyos, C., Estévez-Ayres, I., Delgado Kloos, C., Villena-Román, J.: From MOOCs to SPOCs… and from SPOCs to flipped classroom. In: Lavoué, É., Drachsler, H., Verbert, K., Broisin, J., Pérez-Sanagustín, M. (eds.) EC-TEL 2017. LNCS, vol. 10474, pp. 347–354. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-66610-5_25
14. Uijl, S., Filius, R., Ten Cate, O.: Student interaction in small private online courses. Med. Sci. Educ. **27**(2), 237–242 (2017). https://doi.org/10.1007/s40670-017-0380-x
15. Kloos, C.D., et al.: Boosting interaction with educational technology. In: Proceedings of the IEEE Global Engineering Education Conference 2017, pp. 1763–1767. IEEE (2017)
16. Cheng, H.N.H., Liu, Z., Sun, J., Liu, S., Yang, Z.: Unfolding online learning behavioral patterns and their temporal changes of college students in SPOCs. Interact. Learn. Environ. **25**(2), 176–188 (2017)
17. Jong, J.P.: The effect of a blended collaborative learning environment in a small private online course (SPOC): a comparison with a lecture course. J. Baltic Sci. Educ. **15**(2), 194–203 (2016)
18. Ai, Y., Zhang, S., Wang, M.: Research on the SPOC effective teaching influencing factors and promotion strategies. Vocat. Tech. Educ. **39**(11), 45–49 (2018)
19. Yang, X., Yu, S.: Research on influencing factors of SPOC mixed teaching effect based on ISM model. Mod. Inf. Technol. **3**(2), 185–187 (2019)
20. Liao, H., Liu, W.: An empirical analysis of influencing factors of SPOC effective learning in colleges and universities. E-educ. Res. **38**(5), 64–70 (2017)
21. Cheung, K.S., Lam, J., Im, T., Szeto, R.: Exploring a pedagogy-driven approach to e-course development. In: Proceedings of the IEEE International Workshop on Education Technology and Training, pp. 22–25. IEEE (2008)

22. Lam, J., Cheung, K.S., Yau, J., Seto, W., Im, T.: Students' needs of e-courses as complement to traditional learning: a Japanese studies e-course case. In: Proceeding of the International Symposium on IT in Medicine and Education, pp. 876–880. IEEE (2008)
23. Henri, F.: Computer conferencing and content analysis. In: Kaye, A.R. (ed.) Collaborative Learning Through Computer Conferencing, vol. 90, pp. 117–136. Springer, Heidelberg (1992). https://doi.org/10.1007/978-3-642-77684-7_8
24. Garrison, D.R.: Critical thinking and self-directed learning in adult education: an analysis of responsibility and control issues. Adult Educ. Q. **42**(3), 136–148 (1992)
25. Newman, D.R., Webb, B., Cochrane, C.: A content analysis method to measure critical thinking in face-to-face and computer supported group learning. Interpersonal Comput. Technol. **3**(2), 56–77 (1995)
26. Gunawardena, C.N., Lowe, C.A., Anderson, T.: Interaction analysis of a global online debate and the development of an interaction analysis model. J. Educ. Comput. Res. **17**(4), 397–431 (1997)
27. Xiong, J.: Research on the collaborative knowledge construction under the online learning environment: interactive level, process and situation design. China Educ. Inf. **3**, 7–12 (2019)
28. Moore, M.G.: Three types of interaction. Am. J. Distance Educ. **3**(2), 1–7 (1989)
29. Li, K., Xie, Y.: Qualitative research on the construction and application of online courses in colleges and universities. Open Educ. Res. **15**(1), 15–21 (2011)
30. Fogg, B.: A behavior model for persuasive design. Persuasive Technology Lab, Stanford University (2009)

Enhancing Academic Reading Skills Using a Peer Assessment of Online Collaborative Annotation Approach

Wing Shui Ng[1(✉)], Haoran Xie[2], and Gary Cheng[3]

[1] School of Science and Technology, The Open University of Hong Kong, Ho Man Tin, Kowloon, Hong Kong SAR, China
wsng@ouhk.edu.hk

[2] Department of Computing and Decision Sciences, Lingnan University, Tuen Mun, New Territories, Hong Kong SAR, China
hrxie@ln.edu.hk

[3] Department of Mathematics and Information Technology, The Education University of Hong Kong, Tai Po, New Territories, Hong Kong SAR, China
chengks@eduhk.hk

Abstract. University students must acquire good academic reading skills in undergraduate study. One of the effective skills is to insert annotations in the process of reading. With the advance of web technologies, students can jointly annotate an online article. However, the effectiveness of using online collaborative annotation platforms largely depends on the integration of instructional support. This paper elaborates on the design and effectiveness of using a peer assessment of online collaborative annotation approach in enhancing academic reading skills. Seventy-five first-year undergraduates were invited to annotate an article collaboratively in an online platform. They were requested to carry out peer assessment on the appropriateness of the annotations. Results from the data collected by a questionnaire suggest that the students considered the online platform enhanced collaborative learning, and the peer assessment component helped engage learning. They expressed their preferences to use the online annotation platform collaboratively in their future learning.

Keywords: Peer assessment · Collaborative annotation · Social annotation · Academic reading skills

1 Introduction

Reading skills have been identified as one of the vital competence in the 21st century (P21 2019). When reading an article, readers are required to engage in a cognitive process to interact with the text (Yuksel and Yuksel 2012). The effectiveness of using cognitive processes to obtain the meaning of an article is affected by the nature of the text, the background knowledge of the readers, and the situation the reading occurs in (Hunt 2004). Usually, readers attempt to construct the meaning of the text by using their

L.-K. Lee et al. (Eds.): ICTE 2020, CCIS 1302, pp. 281–292, 2020.
https://doi.org/10.1007/978-981-33-4594-2_24

knowledge of vocabulary and language. In the process, readers from time to time are required to form hypotheses of their understanding of the text and test their predictions (Zhang 2001). In education, the ability to read to understand the meanings conveyed by an article is regarded as a fundamental skill of information literacy (Mizrachi 2015).

Particularly in the academic world, to succeed in undergraduate study, students must acquire a good level of academic reading skills (Hermida 2009; Sheorey and Mokhtari 1994). However, even though academic reading skills are important, it is surprising that reading skills are seldom taught in university courses (Herimida 2009). University lecturers may simply assume that students already acquired these skills before they join the university (Erickson et al. 2006). However, university-level readings materials are much different from that in high school. It has been identified that most first-year university students lack effective academic reading skills (Herimida 2009). In view of the situation, there is a strong need to equip first-year university students with academic reading skills to support their learning in the university. This paper aims to report a study of using a peer assessment of online collaborative annotation to enhance students' academic reading skills.

2 Literature Review

This section, firstly, introduces some common skills to enhance reading comprehension. The skill of making annotations to strengthen reading comprehension is particularly highlighted. At the end of this section, the development and the functions of online collaborative annotation platforms are elaborated.

2.1 Common Skills to Enhance Reading Comprehension

One of the common skills to enhance reading comprehension is to use a concepts map to depict the linkage of concepts or facts in an article. Students, based on their understanding of the article, create a concept map with linkages of concepts. The concept map serves as a visual aid to consolidate their understanding. For example in the study by Tajeddin and Tabatabaei (2016), the students, in the experimental group, were invited to use concept map tools to construct their understanding on the given readings, while students of the control group read the same text without any concept map tools. The results indicated that the experimental group had a better performance in reading comprehension and information retention.

Another method to enhance reading comprehension is to request students to develop a double-entry journal. The journal is a table with two parallel columns. When reading an article, students extract some information, such as a specific term or a concept, from the text and put it in the left column. The students then enter their responses, analyses or comments in the corresponding right column (Hermida 2009). By doing in this way, students can highlight critical information in the article, and interpret the meaning of the extracted information using their own words. This activity allows students to apply high-order cognitive skills to process the information in the academic paper. The double-entry journal can also be served as evidence of reading comprehension for assessment purposes (Amin 2012). Research has suggested that the method of using a double-entry journal is beneficial to reading comprehension (Amin 2012), and it improves students' motivation to read (Hermida 2009).

Another commonly used skill is to insert annotations to specific texts of an academic article in the process of reading (Lu and Deng 2013). Annotation can take many forms and serve many purposes. It involves inserting markers, such as underlying or highlighting texts, to some important information of the original document. Readers may also write some explanatory notes on it. According to Ovsiannikov et al. (1999), annotations help students retain information, reflect on the concepts in the article, clarify their understanding, and share essential information. It serves as a tool to facilitate active reading to connect what was read and what is written (Kawase et al. 2009; Mendenhall and Johnson 2010). By requesting students to insert annotations in the process of reading an article, they are engaged in a deep approach to reading using high-level cognitive processes to explore the meaning of the text and to identify to the interconnection of concepts (Lu and Deng 2013; Su et al. 2010). Research has confirmed that making annotation improves reading comprehension and students' learning (Chen and Chen 2014; Lu and Deng 2013).

2.2 Online Collaborative Annotation

Traditionally, annotations are marked on a printed document. It is normally developed by an individual and mainly for personal future reference. Annotations are some valuable consolidations on the understanding of an academic article after students engaged in a high-level cognitive process. The annotations are not only beneficial to the annotator, but it also helps others if the annotations are in a readable and understandable format. However, even though the annotations are useful to enhance comprehension, the physical restriction of the paper-based approach inevitably limits the purpose of sharing and collaborative learning (Kawase et al. 2009).

With the advance of internet technology, a few online collaborative annotation platforms, also known as social annotation tools, have been developed (Novak et al. 2012). These online systems allow multiple students to create annotations in the same document together synchronously or asynchronously on the web. The annotations are automatically saved in the online system for collaborators to review thereafter.

A major advantage of creating digital annotations in an online platform is the potentials to enhance collaborative learning. Students, reading the same academic article, can interact and discuss together in the platform. The annotations are shared on the web, and it can be accessed without the restrictions of time and physical location (Novak et al. 2012). By annotating an academic article using an online annotation platform, students are engaged in collaborative learning to explore the meaning or discuss the concepts presented in an academic article (Kawase et al. 2009; Luo et al. 2013). By jointly annotating an academic article in an online collaborative annotation platform, it not only facilitates knowledge construction but also enhances learning motivation and develops professional relationships among students (Kawase et al. 2009). Research has suggested that annotation technologies are generally able to promote engagement, improve critical thinking and enhance reading comprehension (Novak et al. 2012).

Although online collaborative annotation platforms generally help improve reading comprehension and reading attitude (Chen and Chen 2014; Lu and Deng 2013; Mendenhall and Johnson 2010), a critical issue is to integrate effective pedagogy to improve learning (Novak et al. 2012; Razon et al. 2012). Technologies are only tools with some affordances to enhance learning. Merely providing an online collaborative platform without integrating appropriate pedagogy cannot automatically elicit the effectiveness. In the study by Razon et al. (2012), the group provided with instructional guidance, including online prompts, guiding inquiries and constant instructor-assistance, demonstrated a better performance in both reading comprehension and engagement. On the contrary, the group without received any instructional support was more passive in reading, and the students were not engaged in the annotation task. They inclined to pay more attention to the technology instead of focusing on the reading materials. Therefore, it is crucial to design effective pedagogy to engage students in collaborative annotation tasks to enhance reading comprehension.

In consideration of the importance of integrating appropriate pedagogy to enhance the effectiveness of using online collaborative annotation platform to improve reading comprehension, the first author applied a peer assessment of online collaborative annotation strategy to engage students in the reading task. The rationales and design of the strategy are elaborated in the following Research Methodology section. This study aimed to explore the research question "What is the impact of applying a peer assessment of online collaborative annotation strategy to improve students' academic reading skills?".

3 Research Methodology

This section introduces the context of the study. It also elaborates the research procedure and the instrument for collecting data.

3.1 Context of the Study

The first author was a teacher trainer of pre-service and in-service teachers. This study was implemented in the Information Technology in Education course taught by the first author. Two classes of first-year students who were taking the course in the Higher Diploma in Early Childhood Education programme (class size of 39 and 36; a total of 75 students, 73 females and two males) were invited to participate in the study. All of them were just graduated from high schools to join the university education. After explained the purposes, all students were willing to sign a consent form to join this study.

3.2 Research Procedure

The students in each class were divided into seven groups, with a group size of about five. A total of 14 groups were formed to participate in the collaborative reading activity. At the beginning of the activity, each student was given a paper-format academic article. They were required to read the article individually. The title of the provided academic article is "When humanity meets technology: rethinking technology in preschool" (Lin 2013). It reports a case study on the rationale and practices of using information technology in pre-primary school education. It is a long research article with 30 pages. Since all the participants were first-year university students with limited experiences of reading an academic article of this length, it was anticipated that they had difficulties to read and hard to identify the critical elements in the article.

After reading the article individually, the members in each group were required to discuss and raise ten questions to ask about the critical information, concepts or findings in the article. They were required to mark the answers to the questions in the paper-format article. Then, all the questions and answers were submitted to the researcher. This experience aimed to engage the students in the traditional method of annotating an article in the process of reading.

After received the questions and answers from all the groups in both classes, the researcher randomly distributed a set of questions to a group in another class. In other words, each group randomly received a set of 10 questions, prepared by a group in another class, about the same article. The arrangement was anonymous. They did not know who the authors of the set of questions were. The researcher then uploaded the article to the online collaborative annotation platform, Open Rev (Fig. 1). It is a free and easily accessible platform for collaborative annotation of scientific publications. The platform allows user to upload papers, form groups for collaborative annotation, and highlight specific texts in a paper to insert annotations. The students were required to respond collaboratively to the given set of questions by highlighting the answers on the article in the online annotation platform. In this way, the students were able to experience the process of online collaborative annotation.

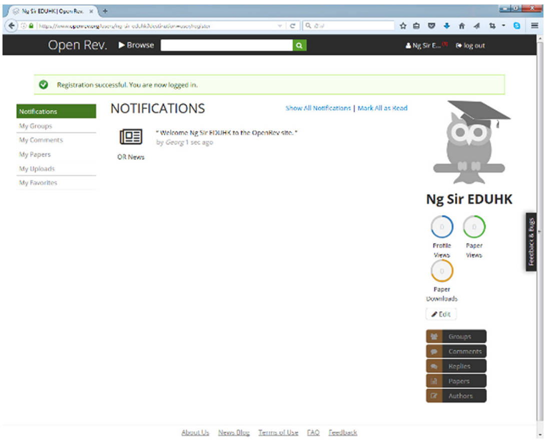


Fig. 1. An online collaborative annotation platform, Open Rev.

As mentioned in the above literature review, simply provide the online annotation platform cannot automatically engaged learners in collaborative learning. To let the students more involved in the online collaborative annotation activity for enhancing reading comprehension, the researcher requested the students in each group to assess the correctness of the answers to the questions raised by their group. They needed to check the highlighted texts in the online article given by the corresponding group in another class. Also, they were required to assess the quality of the questions developed by the group in another class. In the process of peer assessment, they needed to consider whether the questions could address the key points of the article for enhancing reading comprehension and the clarity of the questions.

Research (Ng 2020) has suggested that assessment is a useful tool to make students more engaged in deep learning, and it can transform learners into reflective practitioners. Particularly, peer assessment has been frequently used in higher education for enhancing learning effectiveness (Ng et al. 2018; 2020; Ng and Cheung 2019; Ng 2012; 2013). As indicated by Topping (1998, 2009), the use of peer assessment in teaching and learning promotes reflection and meta-cognition of self-awareness. It also enhances social skills and learning to learn skills. Moreover, students engaged in peer assessment will develop a greater sense of ownership in their learning and a higher motivation to learn. With these underpinning rationales, the researcher deliberately designed the peer assessment activities to engage the students in the online collaborative annotation activity to enhance academic reading skills in this study.

At the end of the activity, the students were invited to fill out a questionnaire to express their opinions. The questionnaire was adapted from that by Ng (2013) and Falchikov (1986). It aimed to explore their views on integrating peer assessment in the online collaborative annotation activity for enhancing academic reading skills. The whole instructional design aimed to improve students' academic reading skills and to enhance their knowledge of applying information technology in preschool education. Their performance in this study did not count into their final grade.

4 Results

Since the participants were the students of the first author who taught the course, all the participants were willing to fill out the questionnaire. The response rate is 100%. The first question asked whether they had any similar experiences of using peer assessment of online collaborative annotation to enhance academic reading skills. The majority (84%) of them gave a negative response. It indicates that the instructional method applied in this study was a new learning experience to them to enhance academic reading skills.

The next section of the questionnaire focused on the peer assessment component. Seven questions listed below asked the impacts brought by the peer assessment strategy.

- The peer assessment component allowed me to learn from my peers.
- The peer assessment component enabled me to spend more time on analysing the contents of the article.
- The peer assessment component enabled me to think critically about the contents of the article.
- The peer assessment component enabled me to have more reflection on my understanding of the article.
- The peer assessment component let me more seriously engage in the learning activity.
- The peer assessment was implemented anonymously. It improved the objectiveness of the assessment.
- I like to learn interactively with peers using peer assessment.

The students were invited to express their opinion using a ten-point scale, with 1 and 10 indicating Strongly Disagree and Strongly Agree respectively. The descriptive statistics are listed in Table 1. The means range from 7.44 to 7.81. To examine whether the means significantly deviate from the neutral stance, the researchers conducted a one-sample t-test on the mid-value (5.5) of the scale. Results show that all the means are significantly above the mid-value (Table 2). It suggests that the peer assessment strategy had significant positive impacts on the students in the online collaborative annotation process.

Table 1. Descriptive statistics of the impacts on integrating peer assessment to enhance academic reading skills

Impact	N	Min	Max	Mean	SD
Learn from peers	75	3	10	7.44	1.726
Spend more time to analyse contents	75	3	10	7.64	1.657
Think more critically on contents	75	4	10	7.44	1.662
More reflection on my understanding	75	3	10	7.47	1.655
Engage more seriously	75	2	10	7.44	1.780
More objective to assess peer	75	4	10	7.81	1.761
Like interactive learning using peer assessment	75	3	10	7.61	1.881
Valid N (listwise)	75				

1 = Strongly Disagree, 10 = Strongly Agree

Table 2. One-sample t-test of the impacts on integrating peer assessment to enhance academic reading skills.

Impact	t	df	Sig. (2-tailed)	Mean difference	95% Confidence interval of the difference	
					Lower	Upper
Learn from peers	9.733	74	.000	1.940	1.54	2.34
Spend more time to analyse contents	11.182	74	.000	2.140	1.76	2.52
Think more critically on contents	10.107	74	.000	1.940	1.56	2.32
More reflection on my understanding	10.292	74	.000	1.967	1.59	2.35
Engage more seriously	9.438	74	.000	1.940	1.53	2.35
More objective to assess peer	11.379	74	.000	2.313	1.91	2.72
Like interactive learning using peer assessment	9.731	74	.000	2.113	1.68	2.55

Test Value = 5.5

The researcher also included the following questions in the questionnaire to investigate their views of using the online collaborative annotation platform to improve academic reading skills.

- The selected article is appropriate for students' learning needs.

- In my previous experience, I used to read individually and annotate on the paper-format articles to learn.
- Compared with reading and annotating paper-format articles individually, the online collaborative annotation approach promotes collaborative learning.
- Compared with reading and annotating paper-format articles individually, the online collaborative annotation approach promotes knowledge sharing.
- Compared with annotating paper-format articles individually, it is more convenient to annotate an article using an online collaborative annotation platform.
- The online collaborative annotation platform is easy to use.
- I will use online collaborative annotation platforms in other courses to improve my learning in the future.

Similarly, the students were invited to express their opinion using a ten-point scale, with 1 and 10 indicating Strongly Disagree and Strongly Agree respectively. The descriptive statistics are listed in Table 3. The means range from 7.33 to 8.24. To examine whether the means significantly deviate from the neutral stance, the researchers conducted a one-sample t-test on the mid-value (5.5) of the scale. Results show that all the means are significantly above the mid-value (Table 4).

Table 3. Descriptive statistics of students' view of online collaborative annotation.

Item	N	Min	Max	Mean	SD
Selected article is appropriate	75	2	10	8.24	1.691
Used to read individually and annotate on paper-format articles	75	1	10	7.88	1.938
Online collaborative annotation approach promotes collaborative learning	75	5	10	8.09	1.526
Online collaborative annotation approach promotes knowledge sharing	75	5	10	7.96	1.520
More convenient to annotate using online collaborative annotation platform	75	4	10	8.08	1.617
Easy to use the online collaborative annotation platform	75	1	10	7.73	1.848
Use online collaborative annotation platforms in future learning	75	4	10	7.33	1.766
Valid N (listwise)	75				

1 = Strongly Disagree, 10 = Strongly Agree

Table 4. One-sample t-test of students' view of online collaborative annotation.

Item	t	df	Sig. (2-tailed)	Mean difference	95% Confidence interval of the difference	
					Lower	Upper
Selected article is appropriate	14.030	74	.000	2.740	2.35	3.13
Used to read individually and annotate on paper-format articles	10.636	74	.000	2.380	1.93	2.83
Online collaborative annotation approach promotes collaborative learning	14.716	74	.000	2.593	2.24	2.94
Online collaborative annotation approach promotes knowledge sharing	14.020	74	.000	2.460	2.11	2.81
More convenient to annotate using online collaborative annotation platform	13.817	74	.000	2.580	2.21	2.95
Easy to use the online collaborative annotation platform	10.467	74	.000	2.233	1.81	2.66
Use online collaborative annotation platforms in future learning	8.993	74	.000	1.833	1.43	2.24

Test Value = 5.5

5 Discussion and Conclusion

Academic reading skills are important for students' learning in the university. However, it is seldom taught in university courses (Herimida 2009). Among the various skills to enhance reading comprehension, making annotations in an article in the process of reading helps engage students in active and deep reading (Lu and Deng 2013; Su et al. 2010). However, annotating an article of paper-format inevitably limits the advantages of sharing annotations for collaborative learning. Some academics (Kawase et al. 2009; Luo et al. 2013) attempted to use online collaborative annotation platforms to engage students in collaborative learning, and they reported positive findings. Actually, the effectiveness of using online collaborative annotation platforms largely depends on the integration of appropriate pedagogies. The study by Razon et al. (2012) clearly stated the importance of instructional guidance to engage students in the online collaborative annotation process. However, there is not much research elaborated on the pedagogies with sound underpinning rationales regarding the application of online collaborative annotation platforms to enhance academic reading skills.

With sound underpinning rationales, the researchers included a peer assessment component to engage students in the online collaborative learning process. Results suggest that the students learned from peers and spent more time to analyse the contents in the article. Their thinking was more critical, and they had more reflection on their understanding of the article. They preferred peer interaction and were more seriously engaged in the learning activity. These findings echo the positive impacts brought by peer assessment in previous studies (Ng 2020; Ng and Cheung 2019; Topping 1998; 2009). It also aligns with Razon et al. (2012) that instructional guidance is critical to engage students in online collaborative annotation activity.

As reflected in the results, the students agreed that the online collaborative annotation approach promoted collaborative learning and knowledge sharing. They found it easy to annotate an article using the online platform. These findings further confirm the effectiveness of using the online annotation platform to enhance collaboration stated in previous studies (Kawase et al. 2009; Luo et al. 2013). As the students expressed that they will use the online collaborative annotation platform in their future learning, and they highly valued the peer assessment component in the process, it suggests that the pedagogy applied in this study had a positive impact on their learning.

Although the results in this study are very favourable, some limitations exist in the research process. First, all the participants were the students taking the same course taught by the first author. In consideration of research ethics, all students should undergo the same teaching experience without any deprivation in learning. Therefore, no control group was included in this intervention study. Moreover, although the students expressed that they will use the online collaboration annotation platform in the future, the researcher did not conduct a longitudinal study to verify whether they used the platform after this study. Therefore, researchers may consider repeating the study with the inclusion of a control group, and to conduct a longitudinal study to explore whether the participants use the online annotation platform in their future learning. It would also be meaningful to study how students interact with their annotations in the online collaborative platform. Notwithstanding the existence of these two limitations, the favourable results obtained from this study suggest the effectiveness of the pedagogy. University lecturers may consider using the peer assessment of online collaborative annotation approach elaborated in this study to engage students and to enhance their academic reading skills.

References

Amin, B.: Double entry journal technique for the students' reading comprehension. Exposure J. **1**(1), 25–34 (2012)

Chen, C.-M., Chen, F.-Y.: Enhancing digital reading performance with a collaborative reading annotation system. Comput. Educ. **77**, 67–81 (2014)

Erickson, B.L., Peters, C.B., Strommer, D.W.: Teaching First-Year College Students. Jossey-Bass, San Francisco (2006)

Falchikov, N.: Product comparisons and process benefits of collaborative peer group and self assessment. Assess. Eval. High. Educ. **11**(2), 146–166 (1986)

Hermida, J.: The importance of teaching academic reading skills in first-year university courses. Int. J. Res. Rev. **3**, 20–30 (2009)
Hunt, R.A.: Reading and writing for real: Why it matters for learning. Atlantic Univ. Teach. Showcase **55**, 137–146 (2004)
Kawase, R., Herder, E., Nejdl, W.: A comparison of paper-based and online annotations in the workplace. In: Cress, U., Dimitrova, V., Specht, M. (eds.) Learning in the Synergy of Multiple Disciplines, pp. 240–253. Springer, France (2009)
Lin, S.S.: When humanity meets technology rethinking technology in preschool. J. Early Child. Educ. Care **10**, 39–68 (2013)
Lu, J., Deng, L.: Examining students' use of online annotation tools in support of argumentative reading. Australas. J. Educ. Technol. **29**(2), 161–171 (2013)
Luo, T., Gao, F., Hoff, K.S.: Examining student participation in three learning activities supported by social annotation tools. J. Educ. Technol. Dev. Exch. **6**(2), 53–68 (2013)
Mendenhall, A., Johnson, T.E.: Fostering the development of critical thinking skills, and reading comprehension of undergraduates using a Web 2.0 tool coupled with a learning system. Interact. Learn. Environ. **18**(3), 263–276 (2010)
Mizrachi, D.: Undergraduates' academic reading format preferences and behaviors. J. Acad. Librarianship **41**, 301–311 (2015)
Ng, W.S.: The impact of peer assessment and feedback strategy in learning computer programming in higher education. Issues Informing Sci. Inf. Technol. **9**, 17–27 (2012)
Ng, W.S.: Students' perspectives on the process and effectiveness of a self- and peer-assessment strategy in learning web design within a Wiki environment. In: Proceedings of the International Conference on Information, Communication Technologies in Education, ICICTE 2013, pp. 257–266. Southampton Solent University, Crete (2013)
Ng, W.S.: Enhancing the quality of educational website design through assessment for learning strategies. In: Khosrow-Pour, M. (ed.) Learning and Performance Assessment: Concepts, Methodologies, Tools, and Applications, pp. 1106–1133. IGI Global, USA (2020)
Ng, W.S., Cheung, K.L.: The effectiveness of peer assessment in the flipped classroom. In: Proceedings of International Conference on Information Communication Technologies in Education ICICTE 2019, pp. 90–100. The International Consortium for Educational Technology, Chania (2019)
Ng, W.S., Xie, H., Wang, F.L.: Enhancing teacher assessment literacy using a blended deep learning approach. In: Cheung, S.K.S., Kwok, L.-f., Kubota, K., Lee, L.-K., Tokito, J. (eds.) ICBL 2018. LNCS, vol. 10949, pp. 203–214. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-94505-7_16
Novak, E., Razzouk, R., Johnson, T.E.: The educational use of social annotation tools in higher education: a literature review. Internet High. Educ. **15**, 39–49 (2012)
Ovsiannikov, I.A., Arbib, M.A., McNeill, T.H.: Annotation technology. Int. J. Hum Comput Stud. **50**(4), 329–362 (1999)
P 21: Frameworks for 21st century learning. retrieved from battelle for kids (2019). http://static.battelleforkids.org/documents/p21/P21_Framework_Brief.pdf
Razon, S., Turner, J., Johnson, T.E., Arsal, G., Tenenbaum, G.: Effects of a collaborative annotation method on students' learning and learning-related motivation and affect. Comput. Hum. Behav. **28**, 350–359 (2012)
Sheorey, R., Mokhtari, K.: The reading habits of developmental college students at different levels of reading proficiency. Read. Improv. **31**(3), 156–166 (1994)
Su, A.Y., Yang, S.J., Hwang, W.-Y., Zhang, J.: A Web 2.0-based collaborative annotation system for enhancing knowledge sharing in collaborative learning environments. Comput. Educ. **55**, 752–766 (2010)
Tajeddin, Z., Tabatabaei, S.: Concept mapping as a reading strategy: Does it scaffold comprehension and recall? Read. Matrix Int. Online J. **16**(1), 194–208 (2016)

Topping, K.: Peer assessment between students in colleges and universities. Rev. Educ. Res. **68**(3), 249–276 (1998)

Topping, K.J.: Peer assessment. Theory Pract. **48**(1), 20–27 (2009)

Yüksel, İ., Yüksel, İ.: Metacognitive awareness of academic reading strategies. Procedia Soc. Behav. Sci. **31**, 894–898 (2012)

Zhang, L.J.: Awareness in reading: EFL students' metacognitive knowledge of reading strategies in an acquisition-poor environment. Lang. Awareness **10**, 268–288 (2001)

Technology and Education

Influence Evaluation of Academic Papers via Citation Characteristics Analysis

Qingyuan Wu[1], Jiahua Hu[2(✉)], Yanghui Rao[3], Fu Lee Wang[4], and Haoran Xie[5]

[1] School of Internet Finance and Information Engineering, Guangdong University of Finance, Guangzhou, China
233137494@qq.com
[2] School of Management, Beijing Normal University, Zhuhai, China
13631265214@163.com
[3] School of Data and Computer Science, Sun Yat-Sen University, Guangzhou, China
raoyangh@mail.sysu.edu.cn
[4] School of Science and Technology, The Open University of Hong Kong, Hong Kong, China
pwang@ouhk.edu.hk
[5] Department of Computing and Decision Sciences, Lingnan University, Hong Kong, China
hrxie2@gmail.com

Abstract. Citation evaluation has become one of the important indexes to evaluate the paper influence. In the academic area, paper influence is quite important to evaluate the author's academic level, in addition to help researchers to identify important academic literatures. The traditional citation evaluation methods are mainly developed based on the citation frequency. Nevertheless, other citation characteristics, such as the nature, depth, and distribution are also valuable. Citation nature includes positive, neutral, and negative citations. The depth can be divided into deep, medium, and shallow citations, which could reveal the correlation and the degree of paper reference between citing and cited papers. The citation distribution contains three types: forepart, middle, and tail citations. The data sample of the empirical study is collected from highly cited papers in Chinese and English core journals. Firstly, by using content analysis methods, we analyze the citation evaluation from three dimensions of Chinese papers. Secondly, we compare the similarities and differences between Chinese and English papers. Thirdly, we show a series of results to verify the accuracy and effectiveness of our method.

Keywords: Citation nature · Citation depth · Citation distribution · Citation evaluation · Content analysis

1 Introduction

Research on the academic value and influence has always been a hot topic in academic researches. Academic papers need to be evaluated in a very subtle way. The traditional method of paper evaluation mainly depends on the citation frequency and citation time. On the basis of the traditional method and the citation content, we propose an in-depth

L.-K. Lee et al. (Eds.): ICTE 2020, CCIS 1302, pp. 295–307, 2020.
https://doi.org/10.1007/978-981-33-4594-2_25

analysis approach in this study, in addition to present multi-type samples in the empirical analysis.

Paper influence analysis is important to evaluate the author's academic level, and to help researchers and learners identify important academic papers for their studies. To evaluate the academic influence of papers, most researches mainly focused on using papers in the same language as the empirical data, which may render the result inaccuracy or incomplete due to the language barrier. For instance, Gross et al. (1927) firstly proposed the Science Citation Index (SCI) and used it to evaluate academic achievements (Garfield 1955). Then, researchers have developed many different citation analysis methods. For example, Hu et al. (2013) considered the citation in two ways, including one paper cites another paper in many positions and a paper cites many papers in one position. Liu (1998) considered the role of critical citation as negative and corrective influence to the cited paper, which is disagree with the basic assumption of citation analysis, that is, "there has a positive relationship between the citation and the quality of the article". Recently, Liu et al. (2013) studied the similarity between cited and citing papers by comparing their abstracts and context contents. Chang (2013) studied the citation status of the target paper over years, and showed that the citation type distribution of the paper was as follows: 'Related literature' was dominant (30.8%); 'evidence' (18.7%) and 'views' (18.7%) were in the second place; 'Terms' (9.2%) and 'background information' (7.2%) were in the third place. Habib and Afzal (2019) proposed a new method of paper recommendation, which extended the traditional bibliographic coupling by using the citation distribution of the logical part of the paper. Wang and Zhang (2019) compared the similarities and differences in citation modes and contents between monographs and papers. Moya-Anegon et al. (2020) proposed advanced indicators that combined citation counts with citation contents, so as to measure the value of cited patents and papers. Jiang and Zhang (2019) used the PageRank algorithm to improve the accuracy of ranking academic papers in the professional Chinese paper database (i.e., CNKI). Geng and Yang (2018) proposed a citation sentiments index, and a way to determine the weight of each citation sentiment. This index could make the evaluation of academic papers more comprehensive. Le (2019) attempted to use the deep learning model to automatically extract reference statements in representative citing papers. Yousif et al. (2019) proposed a multi-task learning model based on recursive neural network, which used task-specific information and shared layers to model citation context and to classify the sentiment/emotion and purpose of citations.

In this paper, we compare the similarities and differences from three dimensions between Chinese and English medical papers, which are citation nature, citation depth, and citation distribution. Considering the time distribution, we select the dataset elaborately and collect the relevant nearly 400 cite papers. Finally, we mainly use the content analysis method to analyze the content and structure. The dataset of this paper was collected from academic papers in two languages, and the dataset is analyzed from three dimensions: the citation nature, the citation depth, and the citation distribution. The content arrangement in the empirical part is as follows:

Firstly, we record the attribute of the dataset, the cited frequency of the dataset, in addition to describe the distribution in three dimensions of each paper. We also observe that some abnormal papers in the data samples, which do not obey the overall distribution.

Considering the negative impact of abnormal data, we filter them out when conducting further studies.

Secondly, we analyze the citation content of the dataset in detail, and compare the results of the similarity and the difference between Chinese and English papers. We also notice that there exists some situations, such as Paper B may cite paper A in different positions or Paper B may cite several papers in the same position. In the above, the citation nature and citation depth are related to the content of the citations, rather than the citation frequency. Regardless of whether the cited behavior occurs once or more, we only record the citation frequency as one.

2 Definition and Methodology

There are two methods to determine the citation nature and the citation depth. One is to infer the author's citation motivation from the content of the paper, and the other is to clarify the author's citation intention by talking with the author. Liu (1998) pointed out that the former was easy to operate but not very accurate, while the latter was highly accurate but difficult to implement. In this study, we refer to use the former method to deduce the author's citation motivation, by employing the Content Analysis method to analyze the full text of the cite papers to infer the citation nature and the citation depth. In the next section, we assume that there are two papers: paper A and paper B, and the relationship is that B cites A or A is citied by B.

2.1 Definition of Citation Nature and Related Concepts

Alvarez and Gómez (2016) first defined sentiment polarity to clarify the author's sentiment when citing papers. In their study, sentiment polarity was divided into three categories: the positive citation, the neutral citation, and the negative citation. This paper adopts this classification system.

The positive citation means the authors of paper B are totally agree with the opinion of paper A, which shows the positive attitudes. In other words, the authors of paper B think paper A is helpful for their research.

The neutral citation means the authors of paper B are not totally agree or basic agree with the opinion of paper A, which shows the neutral attitudes. For instance, the authors of paper B may point out the seldom errors of the cited literature. The other purpose of neutral citation is to lead the argument, and use paper A as background material for readers' reference.

The negative citation means the authors of paper B are totally disagree or basic disagree with the opinion of paper A, which shows the negative attitudes, i.e., the citation's authors deny opinion of the cited literature.

2.2 Definition of Citation Depth and Related Concepts

He (1991) defined the citation depth, which is an intelligence indicator, to reveal the knowledge utilization and the degree of absorption. This article divided the citation depth into three types: the deep citation, the medium citation, and the shallow citation.

The deep citation means that the citation is highly relevant to the cited literature. Particularly, long sentences and contents tend to be quoted, e.g., paper B quotes the core content of paper A and adds innovative research based on the methods of paper A.

The medium citation means that the citation is generally related to the cited literature. For paper A, the sentences and content tend to be quoted, but the quoted length is often shorter than deep citation, i.e., paper B cites non-core or partially core views of paper A and the views are often used as foundation or background. In the medium citation, the reference relationship between papers A and B lies in the middle of the deep citation and the shallow citation.

The shallow citation means that the citation is little related to the cited literature. For paper A, the words tend to be quoted, and the quoted length is the shortest, i.e., paper B only cites non-core views or shallow conclusions of paper A.

2.3 Definition of Citation Distribution and Related Concepts

Citation distribution denotes the location where the cited literature is cited. In our study, we divide the citation distribution into three types: the forepart citation, the middle citation, and the tail citation.

The forepart citation means that the location of the cited literature tends to appear in the former part, normally is the introduction or background part; The middle citation means that the location of the cited literature tends to appear in the middle position which refers to methods, samples, main research processes and so on; The tail citation means that the location of the cited literature tends to appear in the latter position, which refers to the conclusion or discussion part. Note that if paper A is cited in different locations in paper B, only the deepest one is recorded as its citation distribution.

3 Empirical Study

3.1 Data Description

We totally collect 22 highly-cited papers in Chinese and English core medical journals as our corpus. The samples are shown in Table 1.

As the results indicate, the cited counts of these papers are relatively large. Due to the limitations of the space and the time, in the following study, we choose citing paper samples from each cited paper. Particularly, the numbers of samples in Chinese and in English are in proportion to the cited numbers.

3.2 Experimental Results

Analysis of the Citation Nature. In this subsection, we describe the distribution of three dimensions for the collected Chinese and English papers respectively, as shown in Table 2 and Table 3.

Table 1. Cited counts of the 22 papers

Paper ID	Highly-cited paper in Chinese	Cited number	Paper ID	Highly-cited paper in English	Cited number
A1	Introduction to the alkaline single cell gel electrophoresis assay	58	B1	Analysis of skin penetration of phytosphingosine by fluorescence detection and influence of the thermotropic behaviour of DPPC liposomes	10
A2	Applications of micronucleus and tests in genetic toxicology (微核与微核试验在遗传毒理学中的应用)	73	B2	Investigation of genetic susceptibility factors for human longevity-A targeted nonsynonymous SNP study	21
A3	Pollution of PAEs in water and the biodegradations studies in China	90	B3	AMP-activated protein kinase signaling activation by resveratrol modulates amyloid-β peptide metabolism	352
A4	Present contamination status of water organic pollutants and parts of persistent organic pollutants in China	71	B4	Sirt1 improves healthy ageing and protects from metabolic syndrome-associated cancer	371
A5	Three-column classification for tibial plateau fractures	64	B5	Attenuation of age-related changes in mouse neuromuscular synapses by caloric restriction and exercise	177
A6	The updated incidences and mortalities of major cancers in China	160	B6	Arterial stiffness is associated with low thigh muscle mass in middle-aged to elderly men	106

(*continued*)

Table 1. (*continued*)

Paper ID	Highly-cited paper in Chinese	Cited number	Paper ID	Highly-cited paper in English	Cited number
A7	Nasopharyngeal carcinoma incidence and mortality in China	80	B7	Fasting promotes the expression of SIRT1, an NAD + -dependent protein deacetylase, via activation of PPARa in mice	68
A8	Community-based screening for chronic kidney disease among population older than 40 years in Beijing	606	B8	Reduction in the appearance of facial hyperpigmentation after use of moisturizers with a combination of topical niacinamide and N-acetyl glucosamine: results of a randomized, double-blind, vehicle-controlled trial	45
A9	Epidemiologic study of chronic kidney disease in Guangzhou urban area	52	B9	In vitro anti-aging activities of Terminalia chebula gall extract	42
A10	The clinical analysis of repairing the soft tissue defects with transplantation of flap	214	B10	Long-term dietary restriction influences plasma ghrelin and GOAT mRNA level in rats	34
A11	Survey on the pharmacological action of astragalus membranaceus (黄芪药理作用研究综述)	601	–	–	–
A12	Research on the biological activity of naringin and naringenin (柚皮苷及柚皮素的生物活性研究)	112	–	–	–

For the distribution of Chinese papers in these three dimensions (i.e., the citation nature, the citation depth, and the citation distribution), we have the following observations: There are mostly positive citations in the citation nature, mostly medium citations in the citation depth, and mostly forepart citations in the citation distribution.

For the distribution of English papers in these three dimensions, the corresponding observations are as follows: There are mostly positive citations in the citation nature,

Table 2. The distribution of three dimensions of the highly-cited Chinese papers

Paper ID	Citations	Citation nature			Citation depth			Citation distribution		
		Positive	Neutral	Negative	Deep	Medium	Shallow	Forepart	Middle	Nail
A1	14	14	0	0	3	10	1	0	12	2
A2	14	13	1	0	0	12	2	6	3	5
A3	12	11	1	0	1	8	3	10	1	1
A4	12	12	0	0	0	7	5	11	1	0
A5	17	15	1	1	12	4	1	6	3	8
A6	26	26	0	0	6	13	7	21	4	1
A7	21	21	0	0	2	13	6	19	0	2
A8	29	27	2	0	4	18	7	10	1	18
A9	13	11	2	0	3	8	2	4	1	8
A10	2	2	0	0	1	0	1	1	1	0
A11	32	32	0	0	5	19	8	17	0	15
A12	34	34	0	0	6	26	2	27	4	3
Total	226	218	7	1	43	138	45	132	31	63

Table 3. The distribution of three dimensions of the highly-cited English papers

Paper ID	Citations	Citation nature			Citation depth			Citation distribution		
		Positive	Neutral	Negative	Deep	Medium	Shallow	Forepart	Middle	Nail
B1	7	7	0	0	0	3	4	2	4	1
B2	9	5	2	2	0	2	7	2	2	5
B3	10	9	1	0	0	4	6	2	5	3
B4	9	9	0	0	0	2	7	6	2	1
B5	8	8	0	0	0	2	6	3	5	0
B6	9	9	0	0	0	4	5	5	4	0
B7	9	9	0	0	0	3	6	1	6	2
B8	10	10	0	0	0	6	4	3	6	1
B9	7	7	0	0	0	5	2	5	2	0
B10	8	8	0	0	0	5	3	1	4	3
Total	86	81	3	2	0	36	50	30	40	16

mostly shallow citations in the citation depth, and mostly middle citations in the citation distribution.

Furthermore, there is only a few neutral references in the cited data and almost no negative references. For example, there is only one neutral citation in "Applications of micronucleus and tests in genetic toxicology" (paper ID: A2). The citation nature of the above paper is shown in Table 4.

Table 4. The citation nature of paper A2

Year	Citations	Positive	Neutral	Negative
2003	1	1	0	0
2004	2	2	0	0
2006	1	1	0	0
2007	1	1	0	0
2008	2	2	0	0
2009	2	1	1	0
2010	2	2	0	0
2012	2	2	0	0
2014	1	1	0	0
Total	14	13	1	0
Proportion	–	93%	7%	0%

It can be seen that positive citations are much more than neutral citations and negative citations, which indicates that the highly-cited literature about medical has a high recognized degree. Considering the rigorous nature of medical disciplines, most researches are based on long-term practical results, thus negative references are rare. Besides, the collected papers, which were published in well-known medical journals, generally had no academic or presentation errors.

As mentioned above, there is only one neutral citation in the cited data of paper A2. After reading the full text, we found that the authors mostly think the result of the cited paper lacks enough validation in clinical, or it is only a simulation result.

Analysis of Citation Depth. From Table 2 and Table 3, it can be observed that the medium citation ranks the first place in the citation depth. Only a few papers include more deep citations than the medium citation. After reading the cited papers, we found that the medium citations were most of survey papers. The majority of these papers cited a part of the core idea as basis or background.

In the following, we select two papers for further study. The paper IDs are A3 and A4, which both have the highest percentage of medium citations, but paper A4 does not have any deep citations (Fig. 1).

Firstly, it can be observed that the percentage of medium citation times in paper A3 reached 67%, which is significantly higher than the deep citation and the shallow citation. Through reading the cited papers, we found that medium citations and shallow citations, which are more than deep citations, are mostly survey papers. The medium citation can not only improve the non-core or partially core views of the paper but also use those views as the foundation or background. Therefore, medium citations account for the majority of these cited papers.

Secondly, the percentage of shallow citation times in paper A4 reached 42%, but there is no deep citations in this paper. Through reading the cited papers, we found that

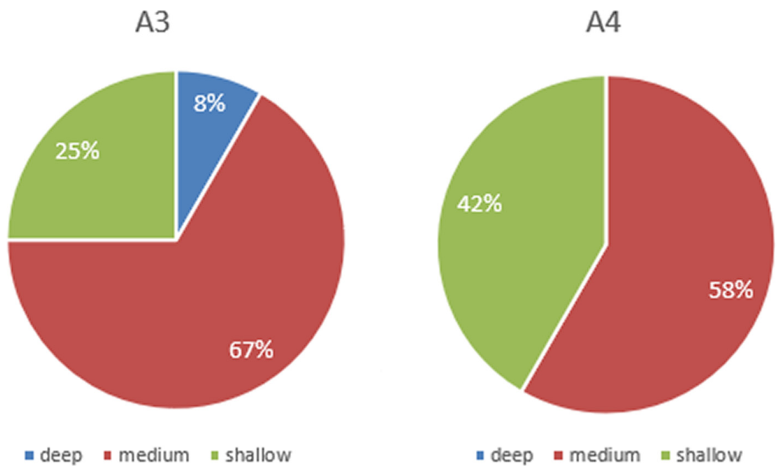


Fig. 1. The citation depth of the selected two papers

the papers in shallow citations are also mostly survey papers. With the development of the medical technology, the citations show that this method had been used in clinic earlier, while authors of paper A4 thought this idea was already out of date and they did not quote the core ideas of the cited literature. So, most of them are shallow citations.

3.3 The Result of Chinese Paper's Citation Distribution

We summarize the results of Chinese papers as follows:

3.3.1 Empirical Results of the Chinese Paper's Citation Nature

As the empirical analysis shows, the percentage of the positive citation number is more than 83%, which indicated that the highly-cited medical papers have a high degree of recognition. As mentioned earlier, the discipline of medicine is rigorous and most of researches are conducted based on the long time practice. Secondly, these papers usually have been published in professional medical journals, which have been strictly reviewed and generally do not allow academic errors. The only negative references that appear in empirical studies is that, the citing authors considered the models in the cited paper is less than ideal, rather than deny the idea of the cited paper.

3.3.2 Empirical Results of the Chinese Paper's Citation Depth

As the empirical analysis shows, the proportion of moderate and deep citations are higher than that of shallow citations. For survey papers, the medium citation is more than the deep citation. For experimental papers, the deep citation is more than the moderate citation, because the core ideas are often referred and the innovative research is carried out on this kind of papers.

3.3.3 Empirical Results of the Chinese Paper's Citation Distribution

The authors conclude the main factors that affect the citation position are the depth of citation and the specific content of the paper. For example, if the content described the research status, there was equally distributed chance on the front location, the middle location, and the tail location. With the depth of citation became deepen, the location of the citation gets closer to the middle of the paper.

4 Comparative Analysis of Highly-Cited Chinese and English Papers

In this section, we compare these highly-cited Chinese and English papers from various aspects.

4.1 Comparative Analysis of Citation Nature

Just like Chinese papers, the English papers had a very high proportion of positive citations, which may be more than 90%. As an illustration, we present the result of one paper named "AMP-activated Protein Kinase Signaling Activation by Resveratrol Modulates Amyloid-β Peptide Metabolism" (paper ID: B3) in Table 5.

Table 5. The nature of citations along the year dimension

Year	Citations	Positive	Neutral	Negative
2018	3	3	0	0
2019	7	6	1	0
Total	10	9	1	0
Proportion	–	90%	10%	0%

There are totally 10 cited papers in the above paper, of which nine are considered as positive citations, and one is neutral citation, which was published by Frolinger et al. (2019). By reading the full content of the cited paper, we find that the authors did not think the method of the cited paper was wrong, but only discovered different research results. Whether it is correct is still to be discussed.

4.2 Comparative Analysis of Citation Depth

For English papers, we can observe that there has no deep citation, and the medium citation, with a percentage of 42%, is less than that of shallow citations. Particularly, there are seven articles having more shallow citations than medium citations, and three articles having more medium citations than shallow citations.

The results of Chinese papers are different. It indicates that most of the cited papers are medium citations. The difference is mainly due to the types of the paper, which can be categorized by subjects and topics. The collected Chinese papers are mostly survey literatures, thus the cite parts used the core idea as background primarily.

4.3 Comparative Analysis of Citation Distribution

We use the full content analysis method to analyze the relationship between citation location and citation depth.

From the cited data of 10 highly-cited English papers, there are 6 papers having the most middle citations and 4 papers having the most forepart citations. It is worth to note that, the citation depth has a strong correlation with the citation distribution. In the statistics of medium citations, the sum of middle citations reached 75%. Furthermore, in the statistics of shallow citations, the sum of forepart and nail citations reached 74%.

For Chinese papers, the percentage of medium citations reached 61%, and forepart citations reached 58%. By using the content analysis, we found that most of medium citations appeared in the introduction or background part.

5 Conclusions

In this part, we summarize our study as follows:

Firstly, we found that the citation nature is mainly positive. Due to peer assessment, negative citations and neutral citations are seldom. Zhao (2010) pointed out that citation evaluation is a part of the scientific evaluation system. In China, researches on the citation evaluation are conducted rapidly, and some evaluation judgement systems have been developed. The traditional citation evaluation used cited quantities to evaluate papers, but it lacks the citation sentiment. For example, suppose there have two papers, both have 20 citing papers. The former paper has 19 positive citations and 1 negative citations, and the latter paper has 15 positive citations and 5 negative citations. Obviously, the citation quality of the former paper should be better than the latter one when considering the sentimental dimension. So when adding sentimental values to traditional citation measurements, it can make the paper's evaluation more accurate. In addition, with the change of time, the frequency of the citation nature maybe changed fundamentally.

Secondly, in terms of the citation evaluation, although deep citations and medium citations are often more important than shallow citations, we should not ignore shallow citations. Zhao (2010) proposed to assign different weights to the citation distribution and the citation depth. In the future, we may need experts to judge the paper's academic status reasonably.

Finally, it can be observed that citation depth and citation content could influence the citation distribution. The deeper the citation, the closer it is to the middle citation.

This study also has some limitations. When using content analysis to identify different citation characteristics, there have some limitations and errors. During the process of judgment, the personal sentiment/emotion is inevitable to make some prejudices. Therefore, in the topic of this academic area, researchers have begun to study the automatic discrimination of citations' content by computer science technologies. However, there have some challenges of evaluating the citation by these methods, like using the computer program to adjust the size of the window of citation context automatically, or identifying the sentimental relationship between the main points of the full text. Despite these challenges, it is necessary to use advanced machine learning algorithms to evaluate citations, which may have a great impact on the development of citation evaluation.

Acknowledgements. The work described in this paper was supported by Philosophy and social science planning project of Guangdong province (GD19YTS02) and a grant from the Research Grants Council of the Hong Kong Special Administrative Region, China (UGC/FDS16/E01/19).

References

Alvarez, M.H., Gómez, J.M.: Survey about citation context analysis: tasks, techniques, and resources. Nat. Lang. Eng. **22**(3), 327–349 (2016)

Chang, Y.W.: The influence of Taylor's paper, question-negotiation and information-seeking in libraries. Inf. Process. Manage. **49**(5), 983–994 (2013)

Frolinger, T., et al.: The gut microbiota composition affects dietary polyphenols-mediated cognitive resilience in mice by modulating the bioavailability of phenolic acids. Sci. Rep. **9**, 3546 (2019)

Garfield, E.: Citation indexes for science: a new dimension in documentation through association of ideas. Science **122**(3159), 108–111 (1955)

Geng, S., Yang, J.: A method to evaluate the academic influence of papers based on citation sentiment. Inf. Stud. Theory Appl. **41**(12), 93–98 (2018). (耿树青,杨建林.基于引用情感的论文学术影响力评价方法研究. 情报理论与实践, 41(12): 93–98, 2018

Gross, P.L.K., Gross, E.M.: College libraries and chemical education. Science **66**(1713), 385–389 (1927)

Habib, R., Afzal, M.T.: Sections-based bibliographic coupling for research paper recommendation. Scientometrics **119**(2), 643–656 (2019). https://doi.org/10.1007/s11192-019-03053-8

He, J.: Citation depth: a study of several relationships in the field of concepts, evaluation indicators and references. Inf. Sci **12**(6), 21–30, 38, 79 (1991). (何佳讯. 引用深度: 概念、评价指标及引用领域若干关系的研究. 情报科学, **12**(6): 21–30, 38, 79 (1991))

Hu, Z., Chen, C., Liu, Z., Hou, H.: From counting references to counting citations: a new way to calculate the total cited times of references. Lib. Inf. Serv. **57**(21), 5–10 (2013). (胡志刚, 陈超美, 刘则渊, 侯海燕. 从基于引文到基于引用——一种统计引文总被引次数的新方法. 图书情报工作, 57(21): 5–10, 2013.)

Jiang, L., Zhang, Q.: Research on the ranking method of academic retrieval results based on citation sentiment interaction. Inf. Stud. Theory Appl. **6**, 172–179 (2019). (姜霖, 张麒麟. 基于引用情感交互的学术检索结果排序方法研究. 情报理论与实践, 6: 172–179, 2019.)

Le, X., et al.: CiteOpinion: Evidence-based evaluation tool for academic contributions of research papers based on citing sentences. J. Data Inf. Sci. **4**(4), 26–41 (2019)

Liu, J.: Evaluation of citation analysis. Inf. Stud. Theory Appl. **21**(2), 87–89 (1998). (刘君玉. 引文分析的评价. 情报理论与实践, 21(2): 87-89, 1998.)

Liu, S., Chen, C.: The differences between latent topics in abstracts and citation contexts of citing papers. J. Am. Soc. Inform. Sci. Technol. **64**(3), 627–639 (2013)

Moya-Anegon, F.D., Lopez-Illescas, C., Guerrero-Bote, V., Moed, H.F.: The citation impact of social sciences and humanities upon patentable technology. Scientometrics (2020). https://doi.org/10.1007/s11192-020-03530-5

Wang, Y., Zhang, C.: Examining similarities and differences of citation patterns between monographs and papers: a case in biology and computer science. Inf. Discov. Deliv. **47**(4), 229–241 (2019)

Yousif, A., Niu, Z., Chambua, J., Khan, Z.Y.: Multi-task learning model based on recurrent convolutional neural networks for citation sentiment and purpose classification. Neurocomputing **335**, 195–205 (2019)

Zhao, Q.: Research on citation character and citation depth in literature. J. Intell. **29**(10), 46–50 (2010). (赵青. 文学学科引用性质与引用深度调查分析. 情报杂志, 29(10): 46–50, 2010.)

Users' Behavioral Intention Toward M-Learning in Tourism English Education: A Case Study of Macao

Xiaohong Wu, Chi Man Tam(✉), and Sha Fang

Faculty of International Tourism and Management, City University of Macau, Avenida Padre Tomás Pereira, Taipa, Macao, Special Administrative Region of China
{xhwu,tamchiman,sfang}@cityu.mo

Abstract. Mobile learning (M-learning) is characterized as a powerful tool of education and learning to facilitate one's learning experiences. Since the flexibility of mobile learning is beneficial to both teachers and learners, it has been widely used in English education. As a tourist destination, tourism English teaching and learning is particularly important in Macao. The study attempts to apply the Technology Acceptance Model (TAM) to explain users' behavioral intention toward M-learning in tourism English education, by extending with two constructs, perceived enjoyment and personal innovativeness. The moderator role of personal innovativeness on the intention to adopt M-learning in tourism English education is also investigated. The online questionnaire is utilized for collecting data across diverse groups of students from different universities in Macao. 252 valid data are analyzed by using partial least square (PLS). The result indicates that perceived usefulness, perceived ease of use, perceived enjoyment, and personal innovativeness significantly affect students' behavioral intention toward M-learning in their tourism English education. However, the moderating effects of personal innovativeness are not supported. The study extends previous research on M-learning adoption, and offers insights into the theoretical and practical development of M-learning in tourism English education in Macao.

Keywords: Mobile learning · Tourism English · Technology Acceptance Model · Perceived enjoyment · Personal innovativeness · Macao

1 Introduction

Mobile phone has been considered as an important learning tool because of its ease of use (Pulla 2017). Mobile learning (M-learning) plays a vital role in education (Thongsri et al. 2018). M-learning allows self-learning and enables learners to have a good control on their learning process. In addition, the flexibility of M-learning offers learners the opportunity to learn freely from the constraints of time and place. Therefore, M-learning is especially suitable for foreign language learning (Liu et al. 2009).

Macao is committed to build a world center of tourism and leisure. Gaming-based tourism industry is the supporting industry for Macao. Tourism English is regarded as

L.-K. Lee et al. (Eds.): ICTE 2020, CCIS 1302, pp. 308–322, 2020.
https://doi.org/10.1007/978-981-33-4594-2_26

one of the most basic abilities for tourism graduates in Macao. Thus, tourism English education is extremely important in Macao. Tourism English is a branch of English education, it can be considered as special-purpose English with specific vocabulary (Tang 2007). Leong and Li (2011) realized the urgency of tourism English education in Macao and proposed that more diversified resources and more innovative learning method should be encouraged to use. M-learning, thus, can be considered as a potential method to enhance the students' learning performance in Macao.

The success of an innovative educational technology highly depends on students' acceptance and adoption. With the popularity of M-learning in education, an increasing research interest in examining the factors that influence its acceptance has been witnessed (Al-Emran et al. 2018). The implementation of M-learning has been widely discussed worldwide, for example, in Mainland China, Thailand, Malaysia, and New Zealand (Tan et al. 2014). However, the studies on M-learning development in Macao remains limited. Therefore, for the successful implementation of M-learning in tourism English education in Macao, it is essential to understand the factors affect students' behavioral intention toward it.

Technology acceptance model (TAM) is one of the most effective and widely used models for predicting new technologies acceptance (Venkatesh 2000). The superior of TAM has been extensively examined in the context of Education (e.g. Leong et al. 2018; Park et al. 2012; Tarhini et al. 2016; Yeou 2016). As a result, TAM is applied as the underlying theoretical model in this study.

Researchers pointed out not only TAM's core variables (perceived ease of use, perceived usefulness), but also other factors should be taken into consideration to predict the acceptance of mobile technology (Yueh et al. 2016). Perceived enjoyment has been confirmed to be closely related to technology acceptance and adoption (Kang et al. 2009). Therefore, perceived enjoyment is considered as a potential driver of students' behavioral intention toward M-learning in the study. Moreover, literatures also suggested that innovative individuals attempt to accept an innovative technology earlier than non-innovators (Agarwal and Prasad 1998). Personal innovativeness positively affects individuals' attitudes toward a new technology (Kaushik et al. 2015). In addition, personal innovativeness also plays a moderator role in some cases. Thus, personal innovativeness has been considered as both an antecedent and a moderator in the present study. Accordingly, this research aims to achieve below objectives:

1. Identify the factors influence undergraduate students' behavioral intention toward M-learning in their tourism English learning in Macao.
2. Investigate the moderating effect of students' personal innovativeness in their tourism English learning.

2 Literature Review

2.1 Mobile Learning

M-learning uses mobile devices for example smartphones, tablet computer, multi-game devices, and personal media players, rather than wired or traditional personal computers (Tan et al. 2014). M-learning is defined as a learning that allows learners to participate in the learning process without time and place constraints (Kukulska-Hulme 2005). Traxler (2009) pointed out that M-learning is a “contextual, individualized and situated learning”. It assists learners’ learning process and allows the learners to collaborate with each other (Al-Emran et al. 2016).

As an innovative education model, M-learning is regarded as a preferred choice in higher education worldwide (Liu 2009). Universities have increased interests and pay more attentions to imply M-learning in teaching (Alrasheedi and Capretz 2015). According to the Educause Center for Applied Research (2017), 70% undergraduate students use mobile phones for their studies.

2.2 Technology Acceptance Model (TAM)

TAM (Davis 1989) is one of the most powerful models to predict uses’ technology acceptance (San-Martín et al. 2015). TAM proposed that users’ behavioral intention to accept a technology is determined by their perceived usefulness and perceived ease of use. According to Davis (1989), perceived usefulness refers to an individual’s perception of enhancing his/her performance and personal well-being by using the technology; perceived ease of use refers to an individual’s perception of cognitive efforts to use the technology. Because of its adaptability and simplicity, TAM has become one of the most widely applied models to measure information technology acceptance (King and He 2006).

Since its introduction, TAM has been widely adopted to examine individual’s technology acceptance and adoption in various context including in M-learning. For example, El-gayar and Moran (2007) utilized TAM to investigate the students’ acceptance of the tablet personal computer (TPC). The result showed that perceived usefulness significantly affects users’ attitude and behavioral intention, and perceived ease of use significantly affects perceived usefulness and attitude. Based on TAM, Wu and Zhang (2014) proposed a unified model to examine users’ continuance intention towards E-Learning 2.0 systems. The result supported that perceived usefulness and perceived ease of use are key predictors of a user’s behavioral intention toward a new technology. Nevertheless, perceived ease of use significantly affects perceived usefulness. Based on the literature, below hypotheses are proposed:

H1. Perceived usefulness has a positive influence on students’ behavioral intention toward M-learning.

H2. Perceived ease of use has a positive influence on students’ behavioral intention toward M-learning.

H3. Perceived ease of use has a positive influence on perceived usefulness.

2.3 Perceived Enjoyment

Perceived enjoyment, an intrinsic factor that leads an individual to accept an information technology, is defined as the extent to which the use of a technology is perceived as enjoyable by itself, regardless of the consequences and compensations that can be achieved (Davis et al. 1992). Liu et al. (2009) proved that the media richness creates playful atmosphere, thus, enhances learner gratification in their e-learning. When a learner finds intrinsically interesting, he/she participates an activity for pleasure and enjoyment rather than for extrinsic rewards. Terzis and Economides (2011) pointed out that perceived enjoyment positively affects confirmation and users' behavioral intention to accept and use of computer-based assessment. In this case, if students perceived enjoyment when they use M-learning, they would be more willing to accept it. Thus, it is proposed as below:

H4: Perceived enjoyment has a positive influence on students' behavioral intention toward M-learning.

2.4 Personal Innovativeness

In the context of M-learning, innovativeness refers to the extent to which an individual accepts a new mobile applications, mobile devices, or games earlier than others (Fatima et al. 2017). The significant influence of individual innovativeness on their behavioral intention has been confirmed in literature. Lu et al. (2011) confirmed that personal innovativeness is a significant predictor of users' intentions to accept mobile internet. Liu et al. (2010) pointed out that personal innovativeness significantly influences students' intentions to use M-learning in Chinese universities. In this case, students with a higher level of innovativeness, especially in a learning context, may have higher desire to accept M-learning. Thus, the hypothesis is proposed:

H5: Personal innovativeness has a positive influence on students' behavioral intention toward accepting M-learning.

Moreover, compared with less innovative users, innovative users have better perceptions of M-learning in terms of its perceived use, ease of use and usage intention (Agarwal and Prasad 1998). It was argued that individual innovativeness may have a moderating effect on the user attitude and behavioral intention of M-learning (Fatima et al. 2017), which leads to the influence of innovativeness on the relationship between user attitude (or its antecedent variables) and behavioral intention for adoption. Therefore, in order to verify the main effect of perceived usefulness, perceived ease of use, and perceived enjoyment on M-learning behavioral intention, it is necessary to study whether there is a moderating effect of individual innovativeness on the relationship between behavioral intention and its antecedent variables. Therefore, the hypotheses are developed as below:

H6a: Personal innovativeness moderates the relationship between perceived usefulness and students' behavioral intention toward M-learning.

H6b: Personal innovativeness moderates the relationship between perceived ease of use and students' behavioral intention toward M-learning.

H6c: Personal innovativeness moderates the relationship between perceived enjoyment and students' behavioral intention toward M-learning.

Based on the above hypotheses, the conceptual model is presented in Fig. 1.

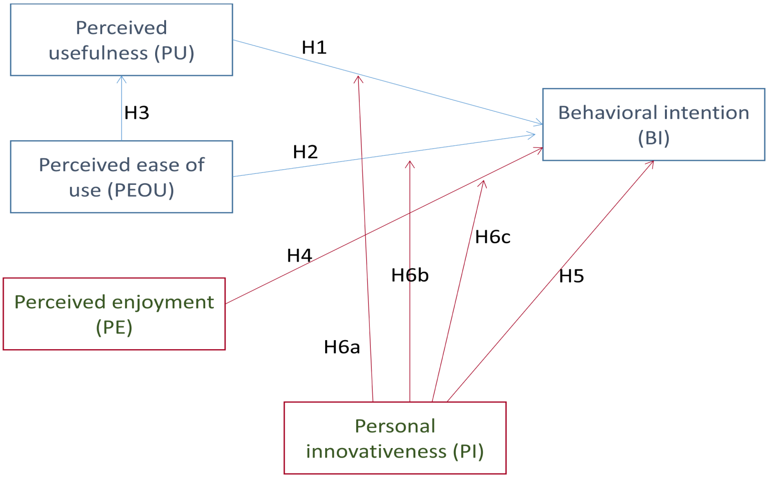


Fig. 1. Conceptual model

3 Research Methodology

3.1 Mobile Application of Tourism English Learning

In this study, a mobile application for tourism English learning was introduced to potential learners (As shown in Fig. 2). The mobile learning application includes various common situations in the travel and tourism industry, such as airports, hotels, car rental, restaurants, shopping malls, etc., so that students can improve their English skills to more immersive situations. Learners can also role-play within mobile learning applications, such as role-playing English conversation for customers or service staff in different situations.

Fig. 2. Mobile application of tourism English learning

3.2 Questionnaire

The prime research question of this study is "What factors affect students' behavior intention to adopt M-learning in tourism English learning?" To exam the theoretical constructs, a questionnaire was designed to determine the factors that influence undergraduate' behavior intention toward M-learning in the context of tourism English learning.

All questionnaire items were based on previous research (see Appendix for details of sources) and were consulted with experts in the field of tourism education to ensure clarity of content, and suitability and format of content. In order to measure perceived usefulness (PU) and perceived ease of use (PEOU), three scale items were used for each construct. In this study, perceived enjoyment (PE) represents the degree of enjoyment using the technology regardless of the expected performance consequences, while personal innovativeness (PI) shows the innovativeness of students in accessing new information such as mobile services and updates. For measuring perceived enjoyment (PE), three scale items have been used, and another three scale items were used for assessing student's personal innovativeness (PI). Finally, behavioral intention (BI) toward mobile-learning in tourism English education adoption has been measured by using three scale items. All survey questions utilized a 5-point Likert scale (1 = strongly disagree, 3 = neutral, 5 = strongly agree). Questions about demographic information were also included in the survey.

As the questionnaire survey was conducted in Macao SAR, the questionnaire was initially compiled in English and then translated into Chinese. Then the questionnaire was translated into English to confirm the equivalence of translation. The validity of the content was reviewed and proofread by a professional translator. In order to test the suitability of the attributes in the questionnaire and correct any errors, we conducted pilot tests on 30 people in Macao SAR to obtain their feedbacks on the questionnaire content.

3.3 Data Collection Method

The target population is limited to Macao undergraduate students who are potential learners of tourism English through M-learning. We used an online survey that was presented in both a web-based format and a mobile device view. The online survey was conducted in June 2020 and 252 valid surveys were collected.

4 Results

After the analysis of the 252 data, the demographic characteristics of the measured students in Macao is shown in Table 1. There are 108 male students and 144 female students, which account for the percentage of 42.9 and 57.1 respectively. As the study is about the university students in Macao, the vast majority age range is between 18 to 25. In addition, the proportion the grade is comparatively equal. And for the English study frequency, more than two third students could study English one to three times per week, one fifth students could study English four to five times per week, about ten

Table 1. Demographic characteristics of the measured students in Macao

		Frequency	Percentage
Gender	Male	108	42.9
	Female	144	57.1
Age	Below 18	1	0.4
	18–25	246	97.6
	26–30	4	1.6
	31–40	1	0.4
Grade	Freshman	44	17.5
	Sophomore	68	27.0
	Junior	99	39.3
	Senior	41	16.2
Frequency of English study	1–3 times per week	157	62.3
	4–5 times per week	52	20.6
	6–7times per week	27	10.7
	More than 8 per week	16	6.4

$N = 252$

percent students could study English six to seven times per week and 6.7% students could study English more than eight times per week.

From Table 2, it can be seen that the values of Cronbach's alpha are all above 0.8. Depending Fraenkel and Wallen (2000), if the Cronbach's alpha greater than 0.7, it means reliability of the scale is acceptable. And all values of average variance extracted (AVE) are above 0.5 (Henseler et al. 2015).

Table 2. Construct reliability and validity

	Cronbach's alpha	rho_A	Composite reliability	AVE
PU	0.897	0.899	0.936	0.830
PEOU	0.807	0.814	0.886	0.722
PE	0.890	0.892	0.932	0.820
PI	0.811	0.815	0.889	0.727
BI	0.871	0.871	0.921	0.795

The Fornell-Larcker Criterion results and Heterotrait Monotrait ratio (HTMT) are showed in Table 3 and Table 4 respectively. It can see that the correlations among all variables and are acceptable.

Table 3. Fornell-Larcker criterion

	BI	PE	PEOU	PI	PI * PE	PI * PEOU	PI * PU	PU
BI	0.892							
PE	0.761	0.906						
PEOU	0.701	0.746	0.850					
PI	0.774	0.637	0.563	0.853				
PI * PE	−0.121	−0.136	−0.076	−0.100	1.000			
PI * PEOU	−0.049	−0.073	−0.069	0.002	0.858	1.000		
PI * PU	−0.075	−0.081	−0.052	−0.061	0.862	0.829	1.000	
PU	0.682	0.711	0.697	0.585	−0.082	−0.050	−0.128	0.911

Table 4. Heterotrait-Monotrait Ratio (HTMT)

	BI	PE	PEOU	PI	PI * PE	PI * PEOU	PI * PU	PU
BI								
PE	0.863							
PEOU	0.834	0.877						
PI	0.920	0.748	0.699					
PI * PE	0.130	0.143	0.083	0.150				
PI * PEOU	0.052	0.078	0.075	0.120	0.858			
PI * PU	0.080	0.086	0.056	0.124	0.862	0.829		
PU	0.771	0.798	0.817	0.685	0.087	0.054	0.134	

Variance inflation factors (VIF) measure the inflation in the variances of the parameter estimates due to collinearities that exist among the predictors. Based on the research of Kline (2010), the value of VIF should below the number of five. Table 5 demonstrates that the values of VIF are between 1.000 to 3.547, which is far below 5. It means that there is no multicollinearity among variables in this study.

According to the results of Table 6 and Fig. 3, perceived usefulness, perceived ease of use, perceived enjoyment, and personal innovativeness have a positive direct influence on students' behavioral intention, and perceived ease of use has a positive impact on students' perceived usefulness. However, the moderate effects have no statistical significance in this study. Table 7 shows the results of all hypotheses testing.

Table 5. Collinearity statistics (VIF)

	VIF
BI1	2.596
BI2	1.990
BI3	2.628
PE * PI	1.000
PE1	2.382
PE2	2.697
PE3	2.858
PEOU * PI	1.000
PEOU1	1.538
PEOU2	2.005
PEOU3	1.935
PI1	1.787
PI2	2.366
PI3	1.702
PU * PI	1.000
PU2	3.498
PU3	2.158
PU1	3.547

Table 6. Mean, standard deviation and P values

	Original sample (O)	Sample mean (M)	Standard deviation (STDEV)	T statistics (|O/STDEV|)	P values
PE -> BI	0.265	0.262	0.065	4.084	**0.000***
PEOU -> BI	0.256	0.260	0.058	4.411	**0.000***
PEOU -> PU	0.697	0.696	0.041	16.912	**0.000***
PI -> BI	0.435	0.435	0.056	7.793	**0.000***
PI * PE -> BI	−0.051	−0.046	0.072	0.701	**0.484**
PI * PEOU -> BI	−0.005	−0.008	0.066	0.078	**0.938**
PI * PU -> BI	0.046	0.046	0.058	0.791	**0.430**
PU -> BI	0.123	0.119	0.059	2.072	**0.039***

Notes: *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$

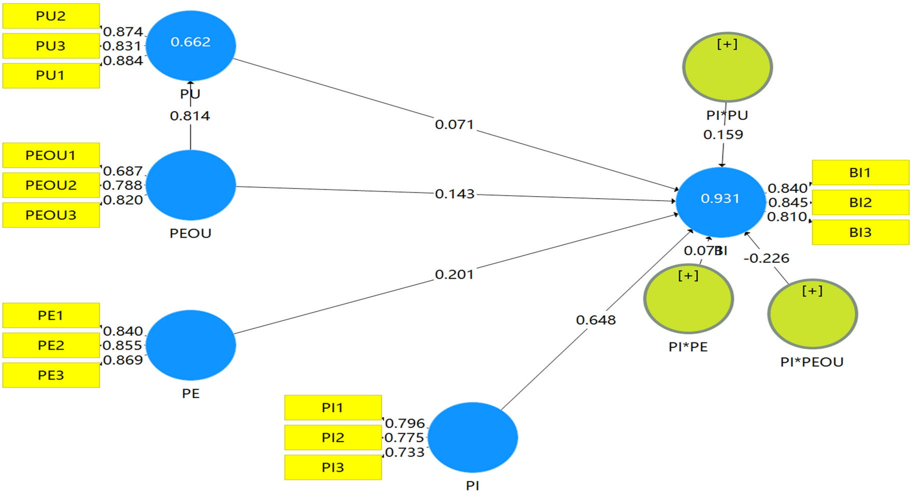


Fig. 3. The results of PLS analysis

Table 7. The results of hypotheses test

FACTORS → Behavioral intention	β-value	p-value	Hypotheses testing
H1: Perceived usefulness → Behavioral intention	0.071	0.000	Accept
H2: Perceived ease of use → Behavioral intention	0.143	0.000	Accept
H3: Perceived ease of use → Perceived usefulness	0.201	0.000	Accept
H4: Perceived enjoyment → Behavioral intention	0.016	0.000	Accept
H5: Personal innovativeness → Behavioral intention	0.648	0.039	Accept
H6: Moderate effort of personal innovativeness between perceived usefulness and behavioral intention	0.159	0.430	Reject
H7: Moerate effort of personal innovativeness between perceived ease of use and behavioral intention	−0.226	0.938	Reject
H8: Moerate effort of Personal innovativeness between perceived enjoyment and behavioral intention	0.071	0.484	Reject

5 Discussion and Conclusion

The study employed and extended TAM to examine undergraduate students' behavioral intention toward M-learning of tourism English. The descriptive statistics revealed that more than two third of respondents intend to use M-learning in tourism English education when it becomes available (73%) and will use it routinely (68%). This indicates that the existing learning methods other than M-learning cannot meet the learning needs of Macao undergraduate student in tourism English. Therefore, it is a good idea for educators to consider changing from the traditional teaching mode relying solely on offline media to the teaching mode combining online and offline.

5.1 Theoretical Implications

This research has certain theoretical implications. Firstly, the study shows that behavioral intention is directly affected by the perceived usefulness and perceived ease of use of the original TAM model. Accordingly, perceived enjoyment is also considered as a relevant determinant of behavior intention using M-learning. In addition, personal innovativeness related to the use of M-learning also plays a role, significantly influencing user behavioral intention.

TAM was initially developed in the context of organization, in order to adopt the model in the context of education, where the end user of the technology are students, to capture the significant result of additional constructs such as personal enjoyment and personal innovativeness is important. In this study, after the TAM model was extended by two constructs, the model explained 74.5% of the variation of M-learning behavioral intention in tourism English learning.

This study demonstrates the positive influence of personal innovativeness on the behavioral intention of M-learning adoption, which makes contributions to the understanding of users' acceptance model. In the past, many new technologies adoption models have used users' perceptions as drivers to determine new technology adoption. However, users may use new technologies because of their personal characteristics. When users decide whether to adopt a new technology that they think might bring convenience and enjoyment, their perception is the crucial factors. But on the other hand, when considering the adoption of new technologies, users show their personal desire to seek new and difference (Hirschman 1980). In this case, personal trait is a driving factor. Therefore, perceptions (such as PU, PEOU and PE) and personal trait (PI), need to be considered simultaneously. The results of the study suggest that researchers should pay attention to perception factors in addition to personal traits when studying new technology acceptance.

The second objective of this study focuses on investigating the moderating effect of personal innovativeness and demonstrates whether higher personal innovativeness in using technology would foster relationship between users' perceptions (PU, PEOU, and PE) and intention to use (BI) M-learning as a tool for their tourism English learning activities. In our conceptual model, PU, PEOU, and PE are considered as antecedent variables of BI. However, if the moderating effect exists, the main effect between the antecedent and dependent variables become ambiguous. It is not appropriate to use these antecedent variables alone to predict BI because the effect also depends on the degree of influence of the moderator variable. From the data analysis in section four, the significant moderating effect of PI on the relationships developed in the related hypothesis (H6a, H6b and H6c) were not found. It suggests that PI exert no moderating effect on the relationship between PU, PEOU, as well as PE, and BI toward adopting M-learning. Since PU $\rightarrow$ BI (H1), PEOU $\rightarrow$ BI (H2), and PE $\rightarrow$ BI (H4) are supported, this study concludes that PU, PEOU, and PE have a positive impact on the behavioral intention of M-learning adoption.

5.2 Practical Implication

By extending the initial TAM model proposed by Davis (1989), in this study, personal innovativeness is the factor that can best predict the behavior intention of M-learning adoption. The high coefficient value ($\beta = 0.648$) indicates that the users' personal trait of innovativeness has a significant impact on their intentions to adopt M-learning. Given that the significant predictive effect of personal innovativeness on behavioral intentions, this finding suggests that marketers of M-learning application should enhance the new experiences offered by this learning tool. The focus can be to provide more opportunities, such as adding auxiliary learning APP on the original tourism English textbooks, so that users including instructors and learners can experience this new type of teaching and learning, and let them show their innovation ability.

Perceived enjoyment is the second most important indicator in the study. In this regard, it would be useful to implement fun and playful elements in M-learning environment. In addition, Cheng (2015) suggested that navigation and convenience will positively affect the perceived enjoyment of M-learning. Since navigation flexibility is considered to be a determinant of users' perceived enjoyment for interactive media, the study suggests that developers of M-learning applications can focus on the flexibility of navigation, such as improving screen layout design and using voice recognition input. Cheng (2015) also concluded that convenience in a mobile environment can increase the fun of learning because users have the ability to learn anywhere and anytime. It is suggested that M-learning providers and marketers should enhance learners' perception of the convenience of mobile learning at any time and any place, deliver a sense of pleasure, and thus improve learners' willingness to use mobile learning.

Perceived usefulness has a positive impact on behavioral intention. Mobile learning is integrated as a form of learning to provide a convenient wireless service, so that learners do not have to use wired devices or traditional personal computers (Tan et al. 2014). It is suggested that M-learning APP marketers emphasize that the system provides unique value by providing utilitarian benefits that other e-learning systems, such as computer applications, cannot provide.

In this study, perceived ease of use has a direct positive impact on behavioral intention, while it also has an indirect impact on BI through perceived usefulness. Based on these findings, this study suggests that M-learning providers should try to develop more user-friendly interfaces and design useful and easy-to-use features to induce learners to use M-learning application. It is also essential to ensure that potential users have easy access to resources that enable their use of mobile learning, such as the feature of one-click installation.

5.3 Limitations

The limitations of the study should be addressed. In addition to the four constructs established in this study, there may be more factors that influence the adoption of M-learning. Further study of antecedent variables can expand our understanding of behavioral intention of mobile learning. Since personal innovativeness is the most important factor in predicting M-learning behavioral intention, further study on the antecedent variable of personal innovativeness will help to better understand the adoption of M-learning.

5.4 Conclusion and Further Studies

This study contributes to the understanding of college student's acceptance of M-learning in tourism English education. The study suggests that educators should consider the combination of online and offline teaching modes. The results of the study confirm that perceived enjoyment and personal innovativeness are the decisive factors for students to accept M-learning. The moderating effects of personal innovativeness to PU, PEOU as well as PE and behavioral intention toward M-learning are not supported.

This study only focused on the behavioral intention of mobile learning in the context of tourism English learning, the results may not be generalizable to different areas. Further research should be undertaken to investigate the applicability of similar concepts for more applications.

Appendix

Table A1. Measurement scales and items

Construct and measurement	Sources
Perceived usefulness (PU) PU1 I believe using mobile-learning in tourism English education improve my learning performance PU2 Using mobile-learning in tourism English education enhances my learning effectiveness PU3 Using mobile-learning in tourism English education easily translates the learning material into specific knowledge	Wu and Zhang (2014); Kim et al. (2010)
Perceive ease of use (PEOU) PEOU1 Using mobile-learning in tourism English education would not require a lot of my mental effort PEOU2 My interaction with mobile-learning in tourism English education would be clear and understandable PEOU3 Mobile-learning in tourism English education would be easy to use	Huang et al. (2007)
Perceived enjoyment (PE) PE1 Using mobile-learning in tourism English would make me feel good PE2 Using mobile-learning in tourism English would be interesting PE3 I would have fun using mobile-learning in tourism English	Huang et al. (2007)
Personal innovativeness (PI) PI1 I like to experiment with new information technology PI2 If I heard about a new information technology, I would look for ways to experiment with it PI3 Among my peers, I am usually the first to try out new information technology	Liu et al. (2010)
Intention for adoption (BI) BI1 I intend to use mobile-learning in tourism English education when it becomes available BI2 If I were asked to express my opinion of mobile-learning in tourism English education, I intend to say something favorable BI3 In the future, I intend to use mobile-learning in tourism English education routinely	Huang et al. (2007)

References

Agarwal, R., Prasad, J.: A conceptual and operational definition of personal innovativeness in the domain of information technology. Inf. Syst. Res. **9**(2), 204–215 (1998)

Al-Emran, M., Elsherif, H.M., Shaalan, K.: Investigating attitudes towards the use of mobile learning in higher education. Comput. Hum. Behav. **56**, 93–102 (2016)

Al-Emran, M., Mezhuyev, V., Kamaludin, A.: Students' perceptions towards the integration of knowledge management processes in M-learning systems: a preliminary study. Int. J. Eng. Educ. **34**(2), 371–380 (2018)

Alrasheedi, M., Capretz, L.F.: Determination of critical success factors affecting mobile learning: a meta-analysis approach. Turkish Online J. Educ. Technol. **14**(2), 41–51 (2015)

Davis, F.D., Bagozzi, R.P., Warshaw, P.R.: Extrinsic and intrinsic motivation to use computers in the workplace. J. Appl. Soc. Psychol. **22**(14), 1111–1132 (1992)

Davis, F.D.: Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Q. **13**(3), 319–340 (1989)

Educause Center for Applied Research: ECAR study of undergraduate students and information technology (2017). https://library.educause.edu/resources/2017/10/ecar-study-ofundergraduate-students-and-information-technology-2017

El-gayar, O., Moran, M.: Examining students' acceptance of tablet PC using TAM. Issues Inf. Syst. **8**(1), 167–172 (2007)

Fatima, J.K., Ghandforoush, P., Khan, M., Masico, R.D., Buhalis, D.: Role of innovativeness and self-efficacy in tourism M-learning. Tour. Rev. **72**(3), 344–355 (2017)

Fraenkel, J.R., Wallen, N.E.: How to Design and Evaluate Research in Education, 5th edn. McGraw-Hill Publishing Co, NewYork (2000)

Henseler, J., Ringle, C.M., Sarstedt, M.: A new criterion for assessing discriminant validity in variance-based structural equation modeling. J. Acad. Mark. Sci. **43**(1), 115–135 (2014). https://doi.org/10.1007/s11747-014-0403-8

Hirschman, E.C.: Innovativeness, novelty seeking, and consumer creativity. J. Consum. Res. **7**(3), 283–295 (1980)

Huang, J., Lin, Y., Chuang, S.: Elucidating user behavior of mobile learning: a perspective of the extended technology acceptance model. Electron. Libr. **25**(5), 585–598 (2007)

Kang, Y., Wang, M., Lin, R.: Usability evaluation of e-books. Disp. **30**, 49–52 (2009)

Kaushik, A.K., Agrawal, A.K., Rahman, Z.: Tourist behaviour towards self-service hotel technology adoption: trust and subjective norm as key antecedents. Tour. Manag. Perspect. **16**(2), 278–289 (2015)

Kim, T., Suh, Y.K., Lee, G., Choi, B.G.: Modeling roles of task-technology fit and self-efficacy in hotel employees' usage behaviours of hotel information systems. Int. J. Tour. Res. **12**(6), 709–725 (2010)

King, W.R., He, J.: A meta-analysis of the technology acceptance model. Inf. Manage. **43**(6), 740–755 (2006)

Kline, B.R.: Principles and Practice of Structural Equation Modeling, 3rd edn. The Guilford Press, New York (2010)

Kukulska-Hulme, A.: Mobile usability and user experience. In: Mobile Learning: A Handbook for Educators and Trainers (2005)

Leong, A.M.W., Li, J.X.: Optimization strategies in English teaching for tourism management majors: a case study of Macau university of science and technology. Int. J. Art. Sci. **4**(21), 315–328 (2011)

Leong, L.W., Ibrahim, O., Dalvi-Esfahani, M., Shahbazi, H., Nilashi, M.: The moderating effect of experience on the intention to adopt mobile social network sites for pedagogical purposes: an extension of the technology acceptance model. Educ. Inf. Technol. **23**, 2477–2498 (2018)

Liu, S.H., Liao, H.L., Pratt, J.A.: Impact of media richness and flow on e-learning technology acceptance. Comput. Educ. **52**, 599–607 (2009)

Liu, X.: Applied research on the mobile learning in foreign language. In: Second International Conference on Education Technology and Training (2009)

Liu, Y., Li, H., Carlsson, C.: Factors driving the adoption of M-learning: an empirical study. Comput. Educ. **55**(3), 1211–1219 (2010)

Lu, Y., Cao, Y., Wang, B., Yang, S.: A study on factors that affect users' behavioral intention to transfer usage from the offline to the online channel. Comput. Hum. Behav. **27**(1), 355–364 (2011)

Park, S.Y., Nam, M.W., Cha, S.B.: University students' behavioural intention to use mobile learning: Evaluating the technology acceptance model. Brit. J. Educ. Technol. **43**, 592–605 (2012)

Pulla, S.: Mobile learning and indigenous education in Canada: a synthesis of new ways of learning. Int. J. Mob. Blended Learn. **9**(2), 39–60 (2017)

San-Martín, S., Prodanova, J., Jiménez, N.: The impact of age in the generation of satisfaction and WOM in mobile shopping. J. Retail. Consum. Serv. **23**, 1–5 (2015)

Tan, W.H., Ooi, K.B., Leong, L.Y., Lin, B.: Predicting the drivers of behavioral intention to use mobile learning: a hybrid SEM-neural networks approach. Comput. Hum. Behav. **36**, 198–213 (2014)

Tang, Y.L.: Feedback of students of regular colleges and universities on present college English teaching: an empirical study. J. Chongqing Inst. Technol. (Soc. Sci. ed.) **8**, 170–173 (2007)

Tarhini, A., Hone, K., Liu, X., Tarhini, T.: Examining the moderating effect of individual-level cultural values on users' acceptance of E-learning in developing countries: a structural equation modeling of an extended technology acceptance model. Interact Learn Environ **25**, 1–23 (2016)

Terzis, V., Economides, A.A.: The acceptance and use of computer based assessment. Comput. Educ. **56**(4), 1032–1044 (2011)

Thongsri, N., Shen, L., Bao, Y., Alharbi, I.M.: Integrating UTAUT and UGT to explain behavioural intention to use M-learning. J. Syst. Inf. Technol. **20**(3), 278–297 (2018)

Traxler, J.: Current state of mobile learning. In: Ally, M. (ed.), Mobile Learning: Transforming the Delivery of Education and Training, pp. 9–25. AU Press, Edmonton (2009)

Venkatesh, V.: Determinants of perceived ease of use: integrating control, intrinsic motivation, and emotion into the technology acceptance model. Inf. Syst. Res. **11**, 342–365 (2000)

Wu, B., Zhang, C.: Empirical study on continuance intentions towards E-Learning 2.0 systems. Behav. Inform. Technol. **33**(10), 1027–1038 (2014)

Yeou, M.: An investigation of students' acceptance of Moodle in a blended learning setting using technology acceptance model. J. Educ. Technol. Syst. **44**, 300–318 (2016)

Yueh, H.P., Lu, M.H., Lin, W.: Employees' acceptance of mobile technology in a workplace: an empirical study using SEM and fsQCA. J. Bus. Res. **69**(6), 2318–2324 (2016)

Cheng, Y.-M.: Towards an understanding of the factors affecting m-learning acceptance: roles of technological characteristics and compatibility. Asia Pacific Manag. Rev. **20**, 109–119 (2015)

Leveraging Neural Network-Based Model for Context Classification of Classroom Dialogue Text

Shunwei Lei[1], Jinlin Li[1], Yu Song[2], Yingshan Shen[1](✉), Lap-Kei Lee[3], and Tianyong Hao[1]

[1] School of Computer Science, South China Normal University, Guangzhou, China
{shunwei,lijinlin,shenys,haoty}@m.scnu.edu.cn
[2] School of Education, South China Normal University, Guangzhou, China
sungyuepku@foxmail.com
[3] School of Science and Technology, The Open University of Hong Kong, Hong Kong, China
lklee@ouhk.edu.hk

Abstract. Classroom dialogue is a common strategy for teaching and learning. Technology-assisted classroom dialogue has drawn increasing attentions, where the classification of classroom dialogue is one of active research topics. However, existing studies mainly paid much attention to dialogue manners rather than dialogue contexts. This paper conducts a deep learning-based experiment on the classification of classroom dialogue context in text format. A hybrid neural network-based model namely CNN-BiLSTM-Attention is proposed for context classification of classroom dialogue text. The hybrid model consists of a Convolutional Neural Network (CNN) and a Bidirectional Long Short-Term Memory Network (BiLSTM) by leveraging an attention mechanism. The CNN-BiLSTM-Attention model is able to capture and learn both the local and global features of classroom dialogue texts for learning semantic information of dialogue contexts. To test the effectiveness of the model, an annotated classroom dialogue text dataset is built based on a well-established coding framework through collecting 155 lessons in Chinese language. Compared with eleven baseline methods, including commonly-used machine learning models and deep learning models, the evaluation results demonstrate that the CNN-BiLSTM-Attention model achieves the best performance with an overall F1-score of 0.7006.

Keywords: Classroom dialogue · Context classification · Deep learning · Neural network

1 Introduction

Classroom dialogue is the main form of traditional classroom teaching, which is intricately linked and intertwined with teaching, and it can make pedagogy more effective and improve the teaching quality with less effort in proper use (van der Veen and van Oers 2017). The theoretical foundation of classroom dialogue has been relatively rich

L.-K. Lee et al. (Eds.): ICTE 2020, CCIS 1302, pp. 323–336, 2020.
https://doi.org/10.1007/978-981-33-4594-2_27

due to its basic role in teaching. Through a review of the research, it is found that the structure of the theoretical definition of classroom dialogue has been focused on the role of individual, interactivity, and dialogue content. Howe and Abedin (2013) confirms the first two parts with their definition based on a review across four decades of research, containing the individual and the interactivity, in which communication is regarded as "one individual addresses another individual or individuals and at least one addressed individual replies" in classroom settings and they divide it into teacher-student interaction and student-student interaction according to the role of individual. When a dialogue needs to be effective, the content of dialogue is supposed to be an indispensable factor. Both the knowledge skills and dialogues are static and it's meaningless if treat them separately. What makes sense is how the dialogue acts as mediation or a tool so that knowledge can be collaboratively constructed and individually appropriated, which is reflected in how the dialogue content is constructed (Wells and Mejía-Arauz 2006). The comprehension of the above three factors plays a guiding role in our subsequent research direction.

Universally, the importance of classroom dialogue continues to be recognized time after time. From a holistic perspective, the researches and applications of classroom dialogue have increased dramatically all over the world in recent years (Song et al. 2019). From a specific perspective, various roles of classroom dialogue are still being explored, even in recent researches. Classroom dialogue was confirmed that it had a positive impact on children's emotion understanding (Pons et al. 2019) and students' thinking and learning, particularly on deep processing of knowledge (Howe et al. 2019). Newell (2019) highly acknowledged the role of dialogue in English classrooms to provide students with the opportunity to explore ideas, actively express and reflect.

With the deepening of academic research, the significance of the classification of technology-assisted classroom dialogue has been increasingly important. Song et al. (2019) reviewed the research and analysis in the field of classroom dialogue, and summarized the main areas of current researches into patterns of classroom dialogue, predictions of variability in dialogue, methodology issues and assessment of classroom dialogue. It can be seen that on the basis of previous theoretical researches, the assessment of classroom dialogue become a new crucial issue, which was manually realized initially. With the increase of datasets, the problem of the timeliness and accuracy of the traditional manpower assessment method is becoming increasingly apparent (Wang et al. 2014; Blanchard et al. 2015). Therefore, in recent years, some scholars have conducted researches on computer-assisted dialogue classification.

Wang et al. (2014) leveraged Random Forest to identify three classifications of classroom dialogue: teacher lecturing, whole class discussion and student group work, based on the features of the percentage of teacher talk time, student talk time, overlapping talking time, silence time, and the average volume. There are several limitations in this research. Firstly, they focused on distinguishing whether or not the teacher or student is speaking and recording both the length of individual's talk time, but the content of dialogue which plays a key role in assessment of classroom dialogue was never involved. The classification based on content of dialogue is quite more instructive for teaching than barely based on utterance duration. Secondly, the classification method was a traditional data mining algorithm, which can be improved to a more efficient method. Thirdly, in implementation, teachers were asked to put the recorder into a pouch worn around their neck which may interfere with their normal classes. Besides, the generality of classification was poor since the classification were limited to mathematics lessons and elementary schools. Based on that, Blanchard et al. (2015) proposed a coding framework of lecture, Q&A, reading aloud, supervision/helping, etc., leveraging Logistic Regression based on the features of number of utterances, mean utterance duration, maximum utterance duration, mean rest duration, and maximum rest duration on 21 classroom recording datasets collected from three teachers. Although the classification had been more precise, it still remained at the level of the individual and interactivity, codding for utterance duration rather than the content of dialogue, and the semantic meaning of the dialogue was not understood. In addition, limitations in method and datasets still exist.

Jiang et al. (2018) improved the method, bringing forward a new hybrid method for overlapping speech detection in classroom environment based on the silence distribution and independent component analysis (ICA). The research focused on whether the segment contained overlapping parts and who were the speakers, to classify classroom dialogue into teacher discourse, student discourse, quiet, and discussion. Nevertheless, it was still struggling with who is talking at the mean time, and the dialogue content was not referred to. The dataset was limited and it was only a speech recorded by the authors to simulate the classroom environment. Lugini and Litman (2018) conducted a research on argument component classification for transcribed spoken classroom discussions. However, it performed poorly on their dataset which has 73 transcripts of text-based classroom discussions. Then, a work leveraged Neural Network Models based on lexical features, parse features, structural features and context features to achieve the goal of automatically classifying student utterances into claims, evidence, and warrants (Howe and Abedin 2013). However, there was a limitation that they solely focus on student utterances. In the class, teachers could guide students to engage in co-constructing, which made teachers occupy a dominant position in most classes (Wells and Mejía-Arauz 2005; 2006). In addition, it could not provide direct feedbacks to teachers for improving teaching in a straight way since teacher-student interaction was not involved.

This research focuses on the context classification of classroom dialogue text based on the technology of natural language processing. The classification of classroom dialogue texts is an indispensable part as well as a fundamental step to analyze and evaluate the teachers' experience in classroom dialogues and provide timely feedback to teachers' instructional content. Only by breaking the limitation of the utterance duration analysis, mastering the analysis of the content of classroom dialogue can the analysis be more instructive so as to improve the teaching quality. To that end, this paper carries out an exploration of utilizing and innovating deep learning-based strategies on the classification of classroom dialogue context. Compared with previous studies, we leverage transfer learning based on a pretrained language model, and a model fusion based on multiple neural networks. We propose a neural network-based model called CNN-BiLSTM-Attention, which consists of an embedding layer, CNN layer, BiLSTM layer, attention layer, and output layer. A pretrained language model RoBERTa (Devlin et al. 2018; Liu et al. 2019) is firstly used to encode and represent the classroom dialogue text in numerical vector form. Then, we leverage a Convolutional Neural Network and a Bidirectional Long Short-Term Memory Network to capture and learn both local and global features of classroom dialogue text in CNN layer and BiLSTM layer respectively, and the attention mechanism is deployed in attention layer, extracting the important features within the dialogue text. Finally, output layer makes the final prediction after calculating the probability distribution of each category. Integrating the corresponding advantage of different components, the model is able to learn deep semantic content of the classroom dialogue text, and the experiment results demonstrate that the classification performance of our CNN-BiLSTM-Attention model is better than traditional machine learning algorithms and other neural network-based methods.

The main contributions of this paper are proposing a neural network-based model to learn semantic content of classroom dialogue, and then leveraging the trained model for context classification of classroom dialogue text.

2 Method

2.1 Coding Framework

There are several researches on coding frameworks for classroom dialogue. Hennessy et al. (2016) developed a coding scheme for analyzing classroom dialogue across educational contexts. They found that concentrating on the "communicative act" to explore dialogue between participants was an appropriate level of granularity, and clustering 33 communicative acts were grouped into 8 clusters, including inviting elaboration or reasoning (ask for explanation or justification of another's contribution, ask for explanation or justification, etc.), building on ideas (build on/clarify others' contributions, clarify/elaborate own contribution), etc. Song et al. (2020) conducted a systematic review on the coding frameworks used in the examination of classroom dialogue. They indicate there are six themes that a dialogic framework should encapsulate in its categories, which includes prior knowledge, personal information, analysis, generalization, speculation and uptakes. These studies lay the theoretical foundation for the classification of classroom dialogue. Based on the research of Song et al. (2020), we adopt the classification scheme which consists of seven categories, including prior-known knowledge, analysis, coordination, speculation, uptake, agreement, and querying.

2.2 Model

A hybrid neural network-based model namely CNN-BiLSTM-Attention, is proposed to classify the classroom dialogue text into specific categories based on the semantic content of dialogic text. The model structure is shown in Fig. 1, consisting of five layers: embedding layer, CNN layer, BiLSTM layer, attention layer, and output layer. The embedding layer is designed to mapped the classroom dialogue text into a numerical vector space by distributed representation of word embedding. The CNN layer is used to capture local semantic information of the classroom dialogue text through Convolutional Neural Network (CNN), while the Bi-LSTM layer utilizes Bidirectional Long Short-Term Memory Network (Bi-LSTM) to learn global context information of the text. The attention layer is deployed to assign different weights to words to enhance understanding of semantic content of classroom dialogue text based on the attention mechanism, and to aggregate the representation of informative words that are important to the semantic content of the text to form feature vectors. The output layer eventually uses a Softmax activation function to calculate the probability distribution of each category, and selects the category with the highest probability as the final category.

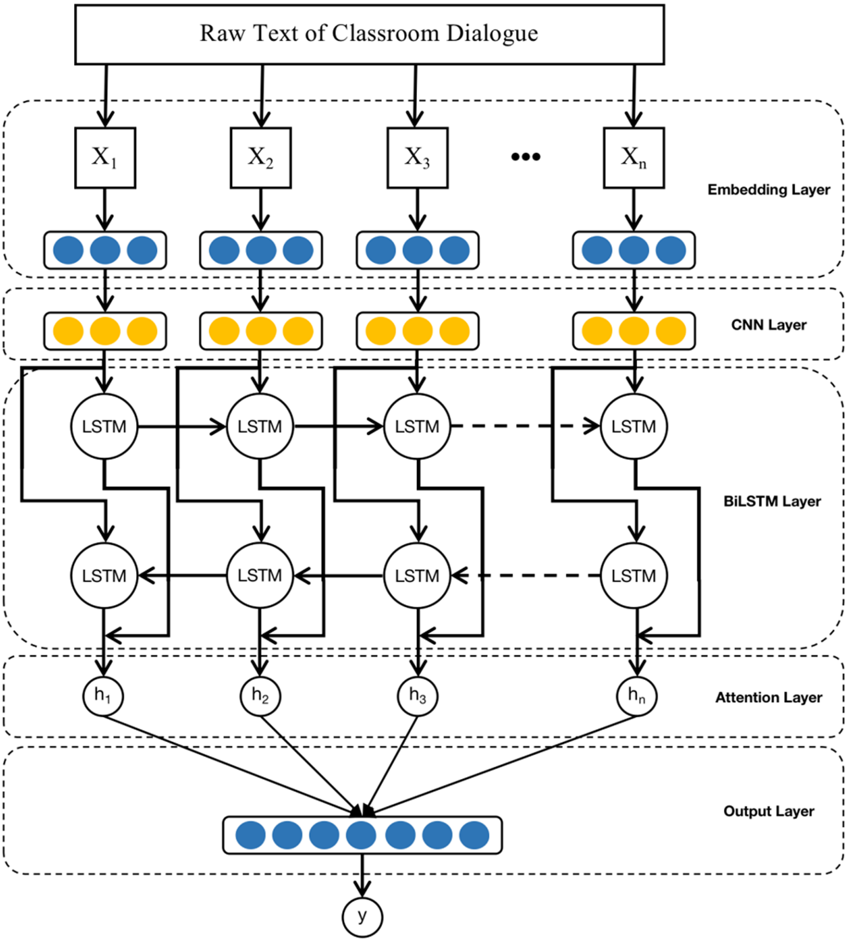


Fig. 1. The structure of the CNN-BiLSTM-Attention model

The first layer in the model is embedding layer, which involves the steps of data preprocessing and word embedding. Since the model cannot process the raw text directly, it is necessary to convert the text into numeric representation. To ensure the independence of input data, we firstly remove all duplicate annotation records. Afterward, we adopt distributed representations of word embedding to encode each sentence with various length of classroom dialogue text into a fixed length vector. Compared to traditional one-hot representations of word embedding, distributed representations address the problems of high-dimensional sparseness and word order loss (Minaee et al. 2020). In detail, we utilize a pretrained language model RoBERTa for transforming each text into numerical representation in the form of low dimensional dense continuous vector. The RoBERTa model is pretrained on large-scale corpora with 35 billion Chinese characters released by CLUE organization, and the model has benefits on pretraining deep bidirectional representations from unlabeled text by jointly conditioning on both left and right context. Thus it is suitable to be applied on our input data.

The second layer of the model is CNN layer, which leverages convolutional neural network to capture local semantic information of the classroom dialogue text (Kim 2014). Two operations including convolution and pooling in this layer are conducted sequentially. We firstly implement a convolution layer with one-dimensional convolution operation. The convolutional filters pass the input text along the input matrix to generate a feature map for representing the local semantics of the dialogic text. After extracting the phrase-level representations, a max-over-time pooling operation over the feature map is conducted to capture the most important feature with the maximum value. Simultaneously, the feature dimension is reduced and the feature representations is compressed or generalized to deal with the overfitting problem.

The third layer of the model is the BiLSTM layer, which utilizes bidirectional long short-term memory network to learn global context information of the classroom dialogue text (Liu et al. 2016). Bi-LSTM is an improved model with bidirectional structure based on Long Short-Term Memory Network, which belongs to a variant of Recurrent Neural Network (RNN). To understand the contextual semantics of long text within classroom dialogue, we utilize Bi-LSTM to extract the interaction information of overall expression. Bi-LSTM has the capability of modeling contextual information dynamically and learning long-term dependencies since it incorporates gate structure to imitate the human memory system to solve the problems of vanishing gradient and exploding gradient in a typical RNN architecture (Graves and Schmidhuber 2005). To capture the bidirectional semantic dependencies, we concatenate hidden state of forward LSTM and of backward LSTM to capture both past and future semantic information, so as to learn the global context semantics comprehensively.

The next layer in the model is attention layer, which is based on attention mechanism (Vaswani et al. 2017) and is deployed to give different focuses on semantic information outputting from the hidden layers within previous layers, and to enhance the understanding for semantic content in dialogic text. In classroom dialogue between teachers and students, there are some keywords partially determining the types of classroom dialogue. For instance, the word "because" usually appears in the analysis process of a teacher or a student. To capture those words that are important in semantic content of the text, we use attention mechanism to calculate corresponding weight of each word, and then weight the sum of the hidden states of all words according to the weights. Each target hidden states and source hidden states of the classroom dialogue text is compared and normalized in the range from 0 to 1 by a Softmax function. Attention weight represents the relevance between each source and target hidden state, and context vector denotes weighted average of source states. The attention vector containing information of current attention decisions is obtained after context vector and target hidden state are concatenated, and projection as well as pass operation are applied to hyperbolic tangent activation function.

The last layer of the model is output layer, which makes the final prediction based on the semantic information learned from previous layers. A Softmax activation function is applied to calculate the probability distribution of each category, in which the category with the highest probability is selected as the final category.

3 Experiments

3.1 Dataset

Due to the lack of publicly available dataset for the context classification of classroom dialogue with corresponding annotations on the existing coding framework, a new dataset is built formally. We firstly collect 155 classroom dialogue videos from elementary school lessons of different subjects and then convert them into text by a semi-automated strategy aided by a computer program with human verification. After that, 15167 sentences from the 155 classroom dialogue texts are manually annotated by two students in Education major. In addition, a professor in the same major, as the administrator, manually reviews the annotations and their differences and the disagreements will be resolved.

The statistical characteristic of the constructed dataset is reported in Table 1. The average count of annotated sentences for each of the seven categories is 2167. In the experiment, the dataset is randomly divided into training dataset and testing dataset, accounting for 80% and 20% of the whole dataset, respectively.

Table 1. The statistics of the classroom dialogue dataset

Categories	Training dataset	Testing dataset	Total
Prior-known knowledge	5714	1429	7143
Analysis	2929	732	3661
Coordination	1270	317	1587
Speculation	375	94	469
Uptake	862	215	1077
Agreement	856	214	1070
Querying	128	32	160
Total	12134	3033	15167

3.2 Baseline Methods

Our CNN-BiLSTM-Attention model is compared with eleven baseline methods in terms of classification performance. The baselines are divided into two types of classification methods. The one type is traditional machine learning algorithms commonly used for text classification, including Logistic Regression (LR), Random Forest (RF), and Support Vector Machine (SVM). In above typical machine learning baselines, the Word2Vec pretrained language model is applied to encode the classroom dialogue text. The other is deep learning classification methods, involving single neural network-based classification models and joint neural networks-based classification models. We adopt two basic artificial neural networks, including Convolutional Neural Network (CNN) and Recurrent Neural Network (RNN). Specifically, the RNN contains two variants, Gated Recurrent Unit (GRU) and Long Short-Term Memory Network (LSTM). Furthermore, we use their bidirectional structure models, consisting of Bidirectional GRU (BiGRU) and Bidirectional LSTM (BiLSTM). Moreover, three joint models are used, including CNN-GRU, CNN-LSTM and CNN-BiLSTM. All of them are combined with two basic types of neural networks, CNN and RNN. CNN-GRU is the combination of CNN and GRU while CNN-LSTM is of CNN and LSTM. CNN-BiLSTM combines the CNN with Bidirectional LSTM. All neural network-based baselines use BERT and RoBERTa to encode the classroom dialogue text.

All baselines are trained for 10 epochs with a batch size of 64 on the same training dataset, using the cross entropy as the loss function and the backpropagation algorithm with Adam optimization method to train the neural networks and to update the model parameters. Within the models containing CNN, the number of convolutional filters is set as 100 and the kernel size is 3. Within the models containing RNN structure, the hidden size is set to 64. ReLu is applied as an activation function.

3.3 Evaluation Metrics

Three commonly evaluation metrics, precision, recall and F1-score, are used to evaluate the effectiveness of the classification models (Olson and Delen 2008). Precision is the proportion of instances with correctly predicted categories over all predicted instances, and recall is the proportion of instances with correctly predicted categories over all instances annotated with the categories. Precision and recall are calculated based on the outcomes of True Positives (TP), False Negatives (FN) and False Positives (FP). For a manual annotation record of classroom dialogue text, the positive category indicates that the actual category annotated by the human coder, while remaining categories, i.e., the negative categories, do not belong to the record. Within the same record, there are two cases for the predicted results with the model. One is that the predicted category is consistent with the target category, and another is that the predicted category is inconsistent with the target category. TP is a measure for the former situation and FN is for the latter. Among those coding records with other categories annotated by the human coder, FP indicates an outcome that the category predicted by the model is not consistent with any one actual category of the coding records. F1-score is the harmonic mean of precision and recall.

3.4 Results

Based on the built dataset, all the classification methods, including traditional machine learning-based methods and deep learning-based methods, run to verify their performance. The performance of our CNN-BiLSTM-Attention model is compared with that of the other 11 baseline methods.

The experiment results, as shown in Table 2, indicate that neural network-based models in overall have better ability to predict the categories of classroom dialogue context than traditional machine learning-based classification models. The LR model with the Word2Vec (Word2Vec + LR) achieves a F1-score of 0.510, which is the best among the traditional machine learning-based models. The SVM model with Word2Vec (Word2Vec + SVM) obtains lower performance than the RF and LR model. Based on the BERT pretrained language model, the GRU model (BERT + GRU) and the LSTM model (BERT + LSTM) obtains similar F1-scores, which is higher than that of the CNN model (BERT + CNN). After applying the bidirectional structure, the F1-socre of BiGRU model (BERT + BiGRU) and BiLSTM model (BERT + BiLSTM) increased from 0.657 to 0.661 (0.6%), and from 0.658 to 0.664 (0.9%), respectively. Furthermore, the joint models CNN-GRU and CNN-LSTM with BERT obtain lower performance than that of the GRU model and the LSTM model. However, the CNN-BiLSTM model with BERT obtains a F1-socre of 0.680, higher than that of the CNN model and the BiLSTM model, with 3.3% improvement. After applying the attention mechanism to the CNN-BiLSTM model (BERT + CNN-BiLSTM-Attention), the performance is further improved, with a F1-socre 0.686, ranking the best among all the models. The experiment has presented that our CNN-BiLSTM-attention model is able to improve the overall classification performance, by introducing the attention mechanism.

Table 2. The performance comparison of the CNN-BiLSTM-Attention model with eleven baseline methods

Methods	Precision	Recall	F1-score
Word2Vec + LR	0.543	0.572	0.510
Word2Vec + RF	0.495	0.530	0.492
Word2Vec + SVM	0.413	0.537	0.424
BERT + CNN	0.657	0.658	0.650
BERT + GRU	0.659	0.658	0.657
BERT + LSTM	0.663	0.670	0.658
BERT + BiGRU	0.662	0.664	0.661
BERT + BiLSTM	0.676	0.664	0.664
BERT + CNN-GRU	0.659	0.648	0.650
BERT + CNN-LSTM	0.647	0.639	0.640
BERT + CNN-BiLSTM	0.688	**0.690**	0.680
BERT + CNN-BiLSTM-Attention	**0.692**	0.685	**0.686**

In addition, to further test the effectiveness of the CNN-BiLSTM-Attention model, we apply another pretrained language model RoBERTa. All the eight neural network-based baseline methods are compared with the CNN-BiLSTM-Attention model. The results, as presented in Table 3, show that the adoption of RoBERTa improves the overall performance of all the neural network-based methods compared with BERT. Among the single neural network-based classification models, the BiLSTM model (RoBERTa + BiLSTM) obtains the best performance with a F1-socre of 0.6748. After combining the CNN model, the joint neural network-based classification model (RoBERTa + CNN-BiLSTM) achieves a F1-socre of 0.6879, with 1.9% improvement. Our CNN-BiLSTM-Attention model (RoBERTa + CNN-BiLSTM-Attention) performs the best, with a precision of 0.7104, a recall of 0.7135, and a F1-score of 0.7006.

Table 3. The performance comparison of the CNN-BiLSTM-Attention model with neural network-based baseline models using the RoBERTa pretrained language model

Methods	Precision	Recall	F1-score
RoBERTa + CNN	0.6611	0.6622	0.6580
RoBERTa + GRU	0.6723	0.6756	0.6705
RoBERTa + LSTM	0.6729	0.6689	0.6582
RoBERTa + BiGRU	0.6752	0.6789	0.6715
RoBERTa + BiLSTM	0.6843	0.6834	0.6748
RoBERTa + CNN-GRU	0.6658	0.6728	0.6660
RoBERTa + CNN-LSTM	0.6738	0.6773	0.6694
RoBERTa + CNN-BiLSTM	0.6907	0.6945	0.6879
RoBERTa + CNN-BiLSTM-Attention	**0.7104**	**0.7135**	**0.7006**

By selecting the RoBERTa + CNN-BiLSTM-Attention model as the best performed method, we further conduct an experiment to view its performance on each category. As shown in Table 4, the results demonstrate that the performance on all the seven categories varies with the imbalanced size of the dataset. The F1-scores of the categories "Prior-known knowledge", "Agreement" and "Analysis" are 0.8394, 0.8350, and 0.7573, respectively, ranking the top three among all the categories. The following categories are "Coordination", "Speculation" and "Uptake" with the F1-scores of 0.6190, 0.5629 and 0.4688, respectively. The "Querying" category has the lowest F1-score of 0.4082. With respect to the Precision metric, our CNN-BiLSTM-Attention model is capable of identifying most of the classroom dialogue text correctly, particularly, on the categories of "Agreement" and "Prior-known knowledge".

3.5 Case Study

As shown in Fig. 2, a sentence "有什么结论?" ('*Is there any conclusion?*') is input into the trained CNN-BiLSTM-Attention model. Based on the training using the training

Table 4. The performance of the CNN-BiLSTM-Attention model on each category

Categories	Precision	Recall	F1-score
Prior-known knowledge	0.8026	**0.8798**	**0.8394**
Analysis	0.6915	0.8369	0.7573
Coordination	0.6603	0.5825	0.6190
Speculation	0.6812	0.4796	0.5629
Uptake	0.6930	0.3543	0.4688
Agreement	**0.8267**	0.8434	0.8350
Querying	0.4762	0.3571	0.4082
Overall	0.7104	0.7135	0.7006

dataset, the model analyzes semantic content of the text and outputs the confidences of all categories. Credibility value ranges from 0 to 1. The label with the highest confidence is selected as the final category of the dialogue sentence. In this case, the label "Coordination" obtains the highest confidence with 0.8386. It indicates that this dialogue sentence is most relevant to the category "Coordination".

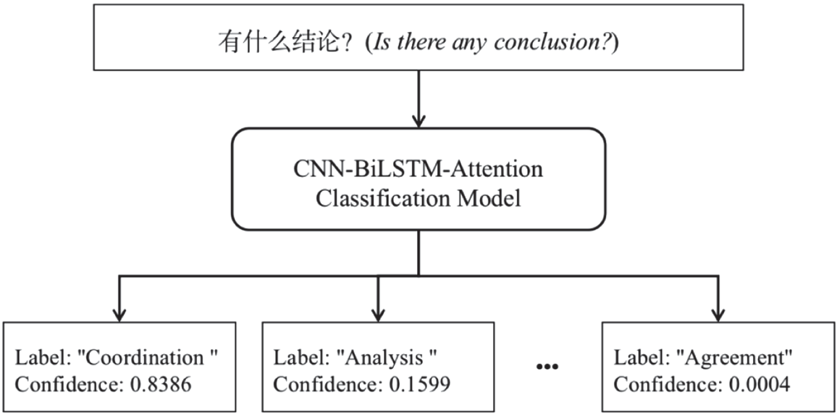


Fig. 2. A case of context classification of classroom dialogue text by CNN-BiLSTM-Attention model

4 Discussion

Compared to the previous studies on the classification of classroom dialogue manner, this paper explores the possibility of context classification of classroom dialogue text based on the deep learning and neural network. To classify the context of classroom dialogue text into appropriate category, it is crucial to understand the semantic content of the text. Because of the complex semantics of the natural language as well as the characteristics of high-dimensionality and sparsity of traditional text representation, the traditional machine learning methods are weak in learning the deep semantic information of the classroom dialogue context. Therefore, we adopt a deep learning approach of natural language processing to capture and learn the semantic information within the classroom dialogue text. Concretely, we propose a hybrid text classification model namely CNN-BiLSTM-Attention, which is comprised of two types of artificial neural networks including Convolutional Neural Network (CNN) and Bidirectional Long Short-Term Memory Network (Bi-LSTM) and is based on the attention mechanism.

Our CNN-BiLSTM-Attention model incorporates five layers, consisting of embedding layer, CNN layer, BiLSTM layer, attention layer, and output layer. In embedding layer, we utilize a pretrained language model RoBERTa to encode the classroom dialogue text into a low-dimensional dense vector space. In terms of CNN layer and BiLSTM layer, we leverage CNN and Bi-LSTM to capture and learn the local semantic information as well as the global contextual semantics of the classroom dialogue text respectively. To enhance the understanding for semantic content in dialogic text, we further deploy the attention mechanism in attention layer. The output layer finally calculates the probability distribution for each category through Softmax activation function, and the category with the highest probability would be selected as the final category. The fusion of different components in various layers strengthens the comprehensive capability of the CNN-BiLSTM-Attention model for learning deep semantic content of classroom dialogue text.

According to the experiment results, the neural network-based models in overall perform better than other traditional machine learning-based classification models such as Logistic Regression (LR), Random Forest (RF), and Support Vector Machine (SVM). It is suggested that neural network-based models are skilled in learning semantic information than traditional machine learning-based models, and it verifies the feasibility of applying deep learning and neural networks for context classification of classroom dialogue text. Compared with other neural network-based models, our CNN-BiLSTM-Attention model obtains the best performance in F1-score when leveraging either BERT or RoBERTa. It reflects the effectiveness of transfer learning based on pretrained language model and of model fusion based on multiple neural networks in context classification task of classroom dialogue text.

Our CNN-BiLSTM-Attention model performs the best on several categories, such as "Prior-known knowledge", "Agreement", "Analysis", but poorly in some categories like "Speculation", "Uptake", and "Querying". In terms of the statistics for our classroom dialogue dataset, there is a circumstance that the category with the most dialogue samples is 43 times more than the category with the least. It is assumed that imbalanced dataset has a certain influence on the model learning on each category. Moreover, there are some major differences between the scores on precision and recall among distinct categories. For instance, the category "Uptake" achieves a precision of 0.693, but obtains a recall of 0.3543. The results show that the model cannot learn enough semantic information from some classroom dialogue text in specific categories. It is critical to improve the model performance in the future work.

5 Conclusion

In this paper, we propose a hybrid neural network-based model namely CNN-BiLSTM-Attention for context classification of classroom dialogue text. Based on a classroom dialogue dataset including 155 lessons in Chinese language, the comparison of the trained model with eleven baseline methods and the experiment results have shown that our model achieves the best performance with an overall F1-score of 0.7006.

References

Blanchard, N., D'Mello, S., Olney, A.M., Nystrand, M.: Automatic Classification of Question & Answer Discourse Segments from Teacher's Speech in Classrooms. International Educational Data Mining Society (2015)

Devlin, J., Chang, M.W., Lee, K., Toutanova, K.: Bert: pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805 (2018)

Graves, A., Schmidhuber, J.: Framewise phoneme classification with bidirectional lstm and other neural network architectures. Neural Netw. **18**, 602–610 (2005)

Hennessy, S., Rojas-Drummond, S., Higham, R., Márquez, A.M., Maine, F., Ríos, R.M., et al.: Developing a coding scheme for analysing classroom dialogue across educational contexts. Learn. Cult. Soc. Interact. **9**, 16–44 (2016)

Howe, C., Abedin, M.: Classroom dialogue: a systematic review across four decades of research. Camb. J. Educ. **43**(3), 325–356 (2013)

Howe, C., Hennessy, S., Mercer, N., Vrikki, M., Wheatley, L.: Teacher-student dialogue during classroom teaching: does it really impact upon student outcomes? J. Learn. Sci. **28**, 1–64 (2019)

Jiang, D., Chen, Y., Garg, A.: A hybrid method for overlapping speech detection in classroom environment. Comput. Appl. Eng. Educ. **26**(1), 171–180 (2018)

Kim, Y.: Convolutional neural networks for sentence classification. In: Conference on Empirical Methods in Natural Language Processing (EMNLP), pp. 1746–1751 (2014)

Liu, P., Qiu, X., Huang, X.: Recurrent neural network for text classification with multi-task learning. arXiv preprint arXiv:1605.05101 (2016)

Liu, Y., et al.: RoBERTa: a robustly optimized BERT pretraining approach. arXiv preprint arXiv:1907.11692 (2019)

Lugini, L., Litman, D.: Argument component classification for classroom discussions. In: 5th Workshop on Argument Mining, Brussels, Belgium, pp. 57–67 (2018)

Minaee, S., Kalchbrenner, N., Cambria, E., Nikzad, N., Chenaghlu, M., Gao, J.: Deep learning based text classification: a comprehensive review. arXiv preprint arXiv:2004.03705 (2020)

Newell, S.: Learning in dialogue: an exploration of talk in an English classroom. Chang. English **26**(4), 357–366 (2019)

Olson, D.L., Delen, D.: Advanced Data Mining Techniques. Springer, Heidelberg (2008). https://doi.org/10.1007/978-3-540-76917-0

Pons, F., Giménez-Dasí, M., Daniel, M.F., Auriac-Slusarczyk, E., Businaro, N., Viana, K.: Impact of a low-cost classroom dialogue-based intervention on preschool children's emotion understanding. Euro. Early Childhood Educ. Res. J. **27**(5), 630–646 (2019)

Song, Y., Chen, X., Hao, T., Liu, Z., Lan, Z.: Exploring two decades of research on classroom dialogue by using bibliometric analysis. Comput. Educ. **137**, 12–31 (2019)

Song, Yu., Hao, T., Liu, Z., Lan, Z.: A systematic review of frameworks for coding towards classroom dialogue. In: Popescu, E., Hao, T., Hsu, T.-C., Xie, H., Temperini, M., Chen, W. (eds.) SETE 2019. LNCS, vol. 11984, pp. 226–236. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-38778-5_25

Van Der Veen, C., Van Oers, B.: Advances in research on classroom dialogue: learning outcomes and assessments (2017)
Vaswani, A., et al.: Attention is all you need. In: Advances in Neural Information Processing Systems, pp. 5998–6008 (2017)
Wang, Z., Pan, X., Miller, K.F., Cortina, K.S.: Automatic classification of activities in classroom discourse. Comput. Educ. **78**, 115–123 (2014)
Wells, G., Mejía-Arauz, R.: Toward dialogue in the classroom: learning and teaching through inquiry. Pap. Trab. Cult. Educ. Desarro. Hum. **1**(4), 1–45 (2005)
Wells, G., Mejía-Arauz, R.: Dialogue in the Classroom. J. Learn. Sci. **15**(3), 379–428 (2006)

Application of Blockchain in Academic Credit Bank System

Ning Yan(✉)

SHCB Management Center, Shanghai Open University, Yangpu, Shanghai, China
ynphd@hotmail.com

Abstract. The essence of blockchain is decentralized and distributed ledger/database technology. Blockchain has the advantages of decentralization, Safety and tamper-proof, traceability, openness and transparency and anonymity. Blockchain technology can effectively deal with the various problems and challenges faced by traditional academic credit bank. In the application scenario of learning achievements deposit, the blockchain technology can avoid import errors or tampering of learning achievements and increase the authority and credibility of the academic credit bank. Automatically executed smart contracts can improve the efficiency of depositing learning achievements into the academic credit bank, reduce labor work and save time. In the application scenario of job application and recruitment, through blockchain, job applicants and employers HR no longer need to re-confirm the certification authority from the certificate issuing institution to achieve the purpose of data confirmation and data verification, which improves the efficiency of job application and recruitment. In the application scenario of academic credit transfer, the use of smart contracts + Oracle for academic credit transfer can effectively integrate students' academic competence information spread across multiple data sources, and improve the efficiency and authority of the academic credit bank system. A three-step strategy can be used to build a blockchain on the academic credit bank system, followed by private chain, alliance chain and public chain. In this paper the account-based blockchain model is preferred rather than the transaction-based blockchain model to be used in the academic credit bank system.

Keywords: Blockchain · Academic credit bank · Smart contract

1 Overview of Blockchain

1.1 Definition of Blockchain

The blockchain is derived from the underlying technology of Bitcoin. In 2008, Satoshi Nakamoto proposed a digital currency called Bitcoin in his paper "Bitcoin: A Peer-to-Peer Electronic Cash System" (Nakamoto 2008). People who do not trust each other can directly pay with Bitcoin, without the coordination of any authoritative intermediary agencies. Ethereum is another well-known blockchain technology application after Bitcoin. The concept of Ethereum was first proposed by Vitalik Buterin after being inspired

L.-K. Lee et al. (Eds.): ICTE 2020, CCIS 1302, pp. 337–348, 2020.
https://doi.org/10.1007/978-981-33-4594-2_28

by Bitcoin in 2013. Ethereum is a public blockchain platform that runs smart contracts, providing a decentralized Ethereum Virtual Machine to process peer-to-peer contracts through its dedicated cryptocurrency Ether (Buterin 2014).

Currently, there are various definitions of blockchain. For example, *The Economist* refers to blockchain as "The Trust Machine" in 2015 (Berkeley 2015). It can be seen that the core problem solved by the blockchain technology is how to allow the two parties in the P2P network to trust each other without credit endorsement, and can conduct transactions with confidence. Many people believe that blockchain technology is one of the main driving forces for the transformation of "Internet of Information" to the "Internet of Value" (Visconti 2019).

This paper summarizes and refines various definitions of blockchain, and concludes that the essence of blockchain is decentralized and distributed ledger/database technology. Blockchain technology is a brand-new distributed infrastructure and computing paradigm that uses blockchain data structures to verify and store data, uses distributed node consensus algorithms to generate and update data, uses cryptographic methods to ensure the security of data transmission and access, and uses smart contracts composed of automated script code to program and manipulate data. Nick Sazb put forward the concept of smart contracts in 1997. His definition of smart contracts is: "a computerized transaction agreement that implements the terms of a contract" (Sazb 1997). The smart contract on the blockchain is an executable program in a sandbox environment. Unlike traditional programs, the smart contract emphasizes transactions, and the smart contract itself is also a program generated by a transaction. The input, output, and state changes of smart contracts all exist in the blockchain, that is, they need to be completed on the basis of the consensus algorithm between nodes. At present, as a core technology of the blockchain, smart contracts have been widely used in Ethereum, Hyperledger Fabric and other influential blockchain projects.

1.2 Advantages of Blockchain

This paper summarizes five major advantages of blockchain as follows:

1. Decentralization: Distributed storage technology can ensure data security and reduce the risk of data loss and hacker attack.
2. Safety and tamper-proof: Unless a malicious node controls more than 51% of the computing power of the entire network, it cannot launch an effective attack; The consensus algorithm ensures that the block data cannot be tampered with, which can effectively eliminate the possibility of possible dark box operations.
3. Traceability: Blocks and timestamps are connected in an orderly manner, ensuring data traceability.
4. Openness and transparency: The data records and updates in the blockchain system are transparent to the nodes of the entire network, which is the foundation of the trustworthiness of the blockchain.
5. Anonymity: The blockchain adopts cryptography technology to solve the problem of trust between nodes. Therefore, data transaction between two nodes can be performed under the condition of anonymity.

1.3 Types of Blockchain

According to the way the verifier participates in the network, blockchains can be divided into three types: public chain, private chain and alliance chain.

The public chain is open to everyone. Users can access the network and blocks anonymously without registration and authorization. Anyone can join and exit the network freely, and participate in accounting and transactions. The public chain is a truly decentralized blockchain. It uses cryptography (asymmetric encryption) algorithm to ensure the security and tamper-proof of transactions. The mutual trust and consensus mechanism are established in a non-secure network environment. In the public chain, the consensus mechanisms are generally Proof of Work (PoW) and Proof of Stake (PoS). Both Bitcoin and Ethereum are public chains.

The alliance chain is limited to members of the alliance, because it only opens all or part of the functions for members, so the read and write permissions and accounting rules on the alliance chain are customized according to the alliance rules. The consensus process on the alliance chain is controlled by pre-selected nodes. In general, it is suitable for B2B scenarios such as inter-institutional transactions, settlement, or clearing. The alliance chain is jointly maintained by participating member institutions and provides a full set of security management functions such as management, authentication, authorization, monitoring, and auditing of participating members.

The private chain is open to individual individuals or entities, and is only used within private organizations, such as companies or organizations. The read and write permissions on the private chain and the permissions to participate in accounting are all set by private organizations. The main value of the private chain is security, traceability, tamper-proof and automatic execution, which is difficult for traditional information systems to achieve simultaneously. Because the private chain has fewer nodes, the transaction speed is faster than the public chain and the alliance chain, and even close to the writing speed of the conventional database.

From the perspective of data structures of blockchain model, blockchain can be divided into transaction-based and account-based blockchains. Bitcoin is a transaction-based model, and Ethereum is an account-based model.

2 Application of Blockchain in Education Field

Firstly, blockchain technology helps to establish a digital certificate issuance, storage and certification system. In the field of blockchain, the term describing the application is "Proof of Existence" (PoE). The basic principle of PoE is to hash the file to be stored, and then store this hash value in the blockchain. Because the blockchain stores all confirmed transactions, and because the hash value is unique, it can prove the existence of a specific file. With the continuous development and popularization of open educational resources such as MOOC, digital certificates based on blockchain technology are ideal tools for storing, sharing and verifying learning achievements. The issuer, receiver, and document signature of the certificate can be stored in a distributed database through the blockchain. This system is decentralized and tamper-proof; any user who has access to the blockchain can perform certificate verification without the involvement of an intermediary party, which reduces the complexity and cost of verification, and the document privacy can be effectively protected. The American company *Learning Machine* has developed *Blockcerts* in conjunction with MIT's Media Lab. *Blockcerts* is an open standard for digital academic certificates based on the Bitcoin blockchain. It provides a decentralized certification system that can be used to issue any type of certificate, including professional certificates, transcripts, credits, or degrees.

Secondly, blockchain technology can promote the transformation of education evaluation methods, giving learners more freedom of learning and resource control. In the process of continuous improvement of education, the evaluation methods of personal achievements will also exhibit diversified development. *Sony Global Education* is developing technology to apply blockchain to the education field to achieve more secure data encryption transmission, especially for students' academic level statistics and measurement data related to exams; to promote the openness and security of academic data Management; through the authorization of students to share test results, to achieve a test result can be evaluated in multiple institutions.

Thirdly, blockchain technology can establish personal electronic portfolios more effectively and securely, and keep records of various learning achievements across different systems and platforms. Learners can collect and combine the academic credits or learning achievements obtained from various educational institutions and apply for certification by educational institutions to obtain course certificates or degree certificates. *OpenBlockchain* is an initiative of the Knowledge Media Institute (KMI) at The Open University in the UK. The "Micro-credentials" and "Open Badges" developed by KMI are supported by *OpenBlockchain.*

Finally, blockchain technology can strengthen the intellectual property protection of digital education resources. On the one hand, digital education resources can facilitate learners to learn and promote education equity; on the other hand, it also brings problems and hidden dangers of intellectual property protection. The traceability of blockchain technology can strengthen the intellectual property protection of digital education resources. For example, *Ledger Journal* is an academic journal about cryptocurrencies and blockchain technology, which requires users to digitally sign their documents with a Bitcoin private key and time stamp them. *Bernstein Technologies* uses blockchain technology for copyright and patent protection, and intellectual property litigation.

3 Overview of Academic Credit Bank

3.1 The Origin of Academic Credit Bank Idea

Lifelong learning is the necessary condition and basic requirement for people to survive in the 21st century. Lifelong learning can occur in any place and environment throughout the life of people. The goal of lifelong learning is to pursue the comprehensive and harmonious development of people. It is expected that people can safely and fully display and express their potential in all stages of their growth. Information technology and information systems have an important support role for lifelong learning.

Some scholars once envisioned a Web system that supports lifelong learning: it can integrate learners' previous learning experience and global resource information, and save all the learning experiences of learners from birth to death which happen in different organizations and environments (Cohn and Hibbitts 2004). The above Web system can be regarded as a Utopian academic credit bank system for lifelong learning. Although the above idea is very attractive, in fact, whether in China or in other countries and regions, individual learning records at different stages are often kept in different organizations and institutions, and in different locations in different data formats. Because various organizations and institutions which have different interests are divided in administration and management, people have to apply for services from different organizations or institutions in order to obtain their previous learning experience data. Therefore, actually people cannot obtain a complete and unified view tool to show their learning experience and learning achievements.

3.2 Overview of Traditional Academic Credit Bank System

Nevertheless, over the years, through the promotion and efforts of the governments and other institutions, some countries and regions, including China and South Korea, have established various academic credit bank systems having different types and levels. Taking Shanghai, China as an example, with the promotion and support of the Shanghai municipal government, Shanghai Academic Credit Bank Management Center for Lifelong Education was established in 2010. It has been running for ten years. This academic credit bank system (www.shcb.org.cn) is connected with multiple information systems, and has established electronic portfolios for millions of people. The various academic records stored in the academic credit bank system include academic education credits, vocational training certificates and community education experiences, etc.

At present, Most successfully built and operated academic credit banking systems are initiated by governments. Through data import or data interface, data from multiple information systems is stored in the central database of academic credit bank system synchronously or asynchronously. This paper refers to the above system as the "traditional academic credit bank system". Generally speaking, traditional academic credit bank systems have three main functions: providing learning achievement deposit and access services, providing academic credit transfer services, and providing learning achievement certification services. The objects of academic credit bank services include both individuals and institutions, that is to say, there are both B2C business and B2B business. People can open an account in the academic credit bank system, access their previous learning achievements, perform academic credit transfer, and authorize third-party institutions to check some or all of their learning achievements or certificates.

From the perspective of participating institutions, the traditional centralized academic credit bank system mainly includes the academic credit bank management center, the academic credit bank branches and the cooperative institutions of the academic credit bank (schools, universities or other certificate issuing institutions).

3.3 Problems and Challenges of Traditional Academic Credit Bank Systems

In summary, traditional academic credit bank systems are facing the following problems and challenges:

1. It is necessary for the traditional academic credit bank center to reach a cooperation agreement with other institutions before it can obtain data through data import or data interface. For the academic credit bank and its cooperative institutions, this method of obtaining data is time-consuming, inefficient, and error-prone.
2. The traditional academic credit bank system is usually designed with the B/S architecture. Both its applications and data are all run or stored on a central server. The centralized application deployment and data storage are prone to problems such as single point of failure, network attacks and data tampering, and the security of data storage and data access is low.
3. It is still difficult and troublesome for the traditional academic credit bank to verify learners' achievements in those institutions that have no cooperative relationship with the academic credit bank, so it is not easy for learners to incorporate a wider range of learning experiences and learning achievements into their academic credit bank accounts, causing their electronic portfolios in the academic credit bank to lack integrity.
4. The update of traditional academic credit bank data records is opaque to other institutions. The data synchronization of the centralized academic credit bank system and the multiple information systems of cooperative institutions is not reliable, and it is prone to data synchronization delay or data inconsistency.
5. Users of the traditional academic credit bank have potential data privacy protection problems. In fact, traditional Internet platforms often have serious data security and privacy protection issues. In 2018, Facebook leaked 50 million user information data is a typical case.

Because of its advantages of decentralized/distributed storage, security and tamper-proof, traceability, transparency, and anonymity, blockchain technology can better address the above problems and challenges. It is even foreseeable that blockchain technology can help to achieve the utopian lifelong learning credit bank mentioned above in a sense.

4 Typical Application Scenarios of Blockchain

4.1 Application Scenarios of Learning Achievements Deposit

In the traditional academic credit bank mode, there are two ways to deposit learning achievements into the academic credit bank. One is the batch deposit of learning achievements by education institutions, and the other is the personal deposit of learning achievements by individuals.

In the mode of batch deposit of learning achievements by education institutions, there may be problems with data synchronization delay or data inconsistency. In the mode of personal deposit of learning achievements by individuals, the problem lies in: on the one hand, there may be errors when manually importing data, and in a few cases, users may falsify their learning records; on the other hand, the staff of the academic credit bank branch need to refer to the transcripts or other certificates to verify the records imported by individuals, and then the staff of the academic credit bank management center will review the learning records before the learning achievements can be finally deposited in the academic credit bank. Both of them a are error-prone labor-consuming and time-consuming. In addition, the possibility of malicious tampering of data by the administrator cannot be completely ruled out.

In the academic credit bank + blockchain mode, once some person or institution submits learning achievements on the blockchain, after the transaction is packaged to form a block, it will be broadcast to all nodes on the blockchain, and it is almost impossible to be tampered with. If the record on the blockchain is found to be wrong, the transaction to modify or delete the record will also be packaged to form a block and broadcast to all nodes on the block chain, which is tamper-proof and traceable. Therefore, the blockchain technology can avoid import errors or tampering of learning achievements and increase the authority and credibility of the academic credit bank. Automatically executed smart contracts can improve the efficiency of depositing learning achievements into the academic credit bank, reduce labor work and save time.

Figure 1 depicts the application of blockchain in the learning achievements deposit scenario of the academic credit bank system. Step 1–3 correspond to the batch deposit mode, and step 4–5 correspond to the individual deposit mode.

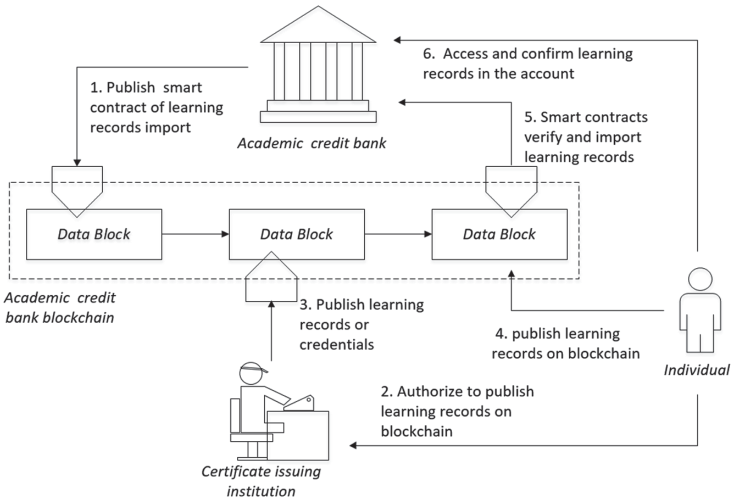


Fig. 1. Blockchain application scenario of learning achievements deposit

4.2 Application Scenarios of Job Application and Recruitment

In the traditional job application and recruitment, the candidate may show his/her academic transcripts and qualification certificates to employers or interviewers to increase the opportunity to get the job. At the same time, it also helps HR find more suitable candidates more effectively and improves recruitment efficiency. However, there are certain deficiencies and limitations in reality:

1. Although job applicants hold academic transcripts and qualification certificates, in fact, the authority of transcripts and certificates still depends on the certificate issuing institution. Despite the certification marks such as seals, in many cases, HR of the potential employer is still insufficient to confirm the authority and authenticity of academic transcripts or qualification certificates.
2. The academic transcripts or qualification certificates of job applicants may be lost or damaged. Individuals need to apply to the certificate issuing institution to obtain a copy of the academic transcript, and it is more difficult to obtain a copy of the certificate. In some cases, the certificate issuing institution is no longer in operation, so people can no longer obtain their lost or damaged certificates.
3. HR of potential employer are generally unable to quickly obtain verification feedback from the certificate issuing institution on the on the academic transcripts or certificates held by job applicants. Availability of timely feedback depends on the certificate issuing institution, so it wastes time for employers and job applicants.

The academic credit bank can effectively deal with the above issues. The learning records in the academic credit bank are imported by the certificate issuing institutions in the credit bank cooperative alliance. These records are authoritative, and in principle, individuals cannot modify his/her learning records in the academic credit bank. So as long as HR of the potential employer is authorized by the job applicant to check his/her credit bank account records, HR can verify the academic transcript and certificates of the job applicant. However, although the number of cooperative institutions of the academic credit bank has been growing, the learning records in individual academic credit bank accounts are still only part of the learning achievements obtained by individuals. It is difficult for the academic credit bank to cover all the learning achievements of users.

If the applicant's transcript or certificate has not been deposited in the academic credit bank, the applicant should still rely on the endorsement of the certificate issuing institution to show the previously obtained certificate to the HR of potential employer. Therefore, although the learning achievements are personally owned, if there is no endorsement by the issuing institution, the owner of the learning achievements cannot authorize his/her learning achievements. How to make the owner of learning achievements endow the authenticity and authority of the learning achievements obtained by oneself is essentially a problem of data authentication. The blockchain can solve the problem of data authentication because of its advantage of tamper-proof and traceability. As long as the certificate issuing institution encrypts the job applicant's learning records and publish them on the blockchain, the job applicant can send the address of learning records on the blockchain to the HR of the potential employer and the academic credit bank. The academic credit bank verifies the authenticity of the job applicant's learning records through smart contracts and deposits it in the applicant's personal account of the academic credit bank. The employer HR can access and verify the job applicant's learning records on the blockchain with the applicant's public key, or the job applicant can authorize the employer HR to access his/her personal account of the academic credit bank to verify his/her learning records. Through blockchain, job applicants and employers HR no longer need to re-confirm the certification authority from the certificate issuing institution to achieve the purpose of data confirmation and data verification, which improves the efficiency of recruitment. The academic credit bank can also obtain credible learning records directly without review, which saves labor and reduces the possibility of data synchronization delay and data inconsistency.

Figure 2 depicts the application of blockchain in the job application and recruitment scenario of the academic credit bank system.

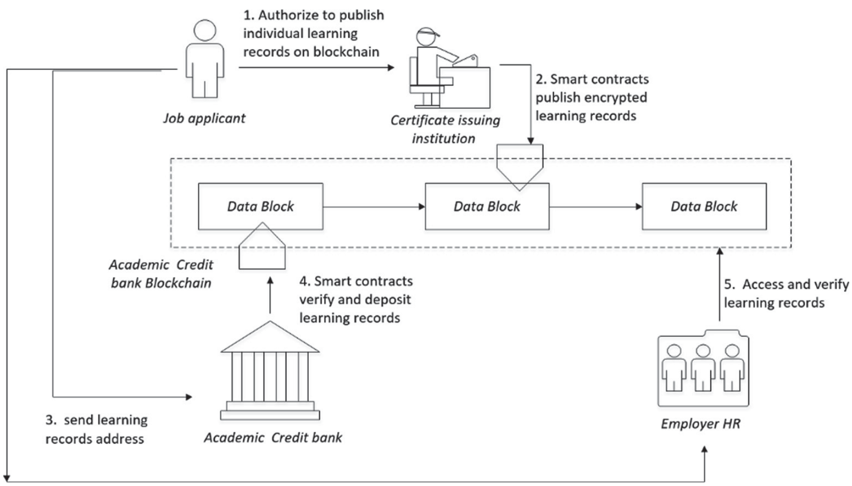


Fig. 2. Blockchain application scenario of job application and recruitment

4.3 Application Scenarios of Academic Credit Transfer

Generally speaking, if students want to transfer their qualification certificates into academic credits, they need to submit the qualification certificates to the academic affairs department of education institutions. then the academic affairs department will convert the students' qualification certificates into the corresponding course credits according to the academic transfer regulation of the education institution. If students transfer, they need to submit the previous transcripts to the transferred education institution, and the transferred education institution will convert the students' previous learning achievements into corresponding course credits according to its academic transfer regulation. There are two deficiencies in the above process. Firstly, students must hold the original transcripts or certificates certified by the certificate issuing institution. If the documents are lost or damaged, they must try to obtain verified copies of the documents from the certificate issuing institution. Students need to spend time and expenses to obtain verified documents. Secondly, transferred education institution also need to spend manpower and time to review the documents submitted by students.

The academic credit bank can effectively deal with the above issues. The cooperative institutions of the academic credit bank publish on the academic credit bank website and regularly maintain credit transfer regulations and corresponding courses/certificates catalogs. Students can refer to the credit conversion rules and courses/certificates catalogs published on the academic credit bank website, and apply for academic credit transfer online, and the academic credit bank cooperative institutions can also accept and review student applications online. All applications and reviews can be conducted online, which not only improves work efficiency and reduces costs, but also reduces the risk of receiving fake documents for education institutions.

However, years of practice have shown that in the real academic credit transfer scenario, the traditional academic credit bank still have the following deficiencies:

1. Actually it is the cooperative institution that can finally decide whether students can perform some academic credit transfer, and the academic credit bank only provide a transfer platform for students and institutions. The credit transfer regulations published on the academic credit bank website are just nominal rules for determining whether credit transfer can be performed, not all. In other words, the actual credit transfer rules also contain many "hidden rules". For example, due to the fact that universities are divided into different levels, most universities tend to transfer credits from universities at the same level or higher than their own level, and do not want to transfer academic credits from universities below their level. This "hidden rule" is usually not included in the published academic credit transfer regulations. Generally speaking, if students need to transfer academic credits from low-level universities to high-level universities, they often need to provide additional information to prove their ability.
2. If students want to transfer academic credits obtained from non-academic credit bank cooperative institutions, they still need to transfer academic credits in the traditional way. Or it is necessary to deposit the original learning achievements in the academic credit bank after verification, then apply for academic credit transfer to the transferred education institution through the academic credit bank. The above process is undoubtedly cumbersome and inefficient.

In the application of blockchain to the academic credit bank, smart contracts + Oracle can be used to deal with the issues in the traditional academic credit bank mentioned above. The function of Oracle is to write external information into the blockchain to complete the data communication between the blockchain and the real world. It allows certain smart contracts to react to the uncertain external world. It is the only way for smart contracts to interact with the outside world, and it is also the interface for the blockchain to interact with the real world.

The academic credit bank and cooperative institutions can write credit transfer regulations and corresponding courses/certificates catalogs into smart contracts and publish them on the blockchain. The Oracle can read some data that the student can prove his/her academic competence off blockchain. For example, this data may be a recommendation letter written by a professor. The Oracle sends the message to the smart contract, and under the manual intervention of the arbitrator (usually the academic administrator), the academic credit transfer is finally completed. Obviously, the use of smart contracts + Oracle for academic credit transfer can effectively integrate students' academic competence information spread across multiple data sources, and improve the efficiency and authority of the academic credit bank system.

Figure 3 depicts the application of blockchain in the academic credit transfer scenario in the academic credit bank system.

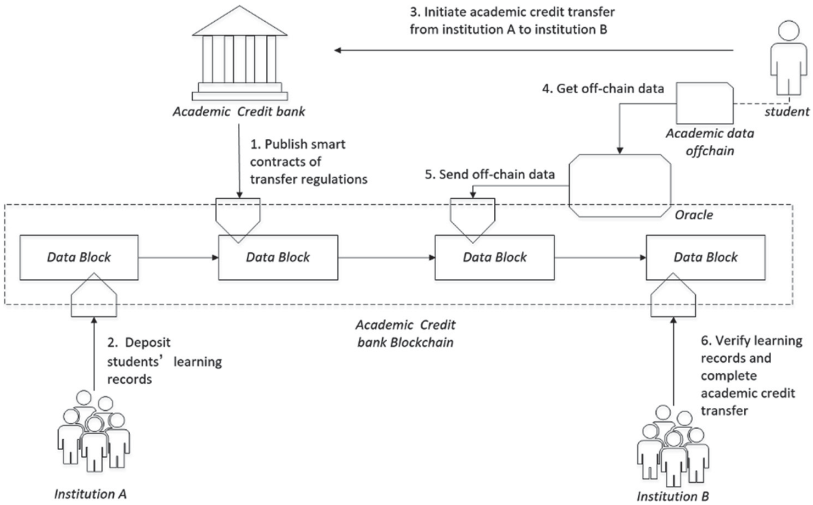


Fig. 3. Blockchain application scenario of academic credit transfer

5 Conclusions and Outlook

Blockchain has the advantages of decentralization, security and tamper-proof, traceability, openness, transparency and anonymity, which can effectively solve the problems and challenges faced by the traditional academic credit bank. In typical application scenarios such as learning achievements deposit, recruitment and job application, and academic credit transfer, the academic credit bank + blockchain mode can save labor and time, reduce the risk of data inconsistency and falsified data, and improve the authority and availability of the academic credit bank. To build a blockchain on the academic credit bank system, it is envisaged in this paper that 3 steps implementation method can be adopted as follows:

1. The internal nodes of the academic credit bank (including the center and branches) use a private chain for operation, which requires authorization to join and access. Because the number of this private chain nodes is relatively small, data security and data upload speed are guaranteed.
2. As a node on the alliance chain, the academic credit bank private chain connects with cooperative institutions nodes. The consensus process on the alliance chain is controlled by the academic credit bank, and the blockchain rules and regulations are jointly formulated by the academic credit bank and cooperative institutions.
3. As a node on the public chain, the academic credit bank alliance chain with other institutions and individuals. Any institution or individual can freely join or exit the academic credit bank public chain.

In addition, the account-based blockchain model is preferred here rather than the transaction-based blockchain model to be used in the academic credit bank. Firstly, transaction-based blockchains, such as Bitcoin, are designed for payment and transfer.

There is no account in the Bitcoin blockchain system, but only the address of the wallet, and a person can have multiple wallet addresses. UTXO is the core idea of Bitcoin, and there are multiple UTXOs in each wallet address. One person's money is the sum of the UTXOs in all these addresses. This design makes the "double spending" in the payment easier to verify. But for learning records, there is no "double spending" problem, academic credits can be used repeatedly. Secondly, for an academic credit bank user, it is more reasonable to have an academic credit account with various learning achievements instead of having multiple academic credit addresses. Thirdly, if the account-based transaction blockchain model is adopted, smart contracts can be applied. Smart contracts would be widely used in multiple application scenarios of the academic credit bank + blockchain, without manual intervention, which can improve work efficiency and reduce execution costs.

Acknowledgements. The author would like to thank the ICTE2020 organizing committee for providing the valuable opportunity, and Pei Hua Education Foundation for the financial sponsorship.

References

Berkeley, J.: The Trust Machine. The Economist 31 Oct 2015 edn (2015). https://www.economist.com/leaders/2015/10/31/the-trust-machine. Accessed 30 Aug 2020

Buterin, V.: Ethereum: A Next-Generation Cryptocurrency and Decentralized Application Platform (2014). https://bitcoinmagazine.com/articles/ethereum-next-generation-cryptocurrency-decentralized-application-platform-1390528211. Accessed 30 Aug 2020

Cohn, E., Hibbitts, B.: Beyond the Electronic Portfolio: A Lifetime Personal Web Space (2004). https://er.educause.edu/articles/2004/1/beyond-the-electronic-portfolio-a-lifetime-personal-web-space. Accessed 30 Aug 2020

Nakamoto, S.: Bitcoin: A Peer-to-Peer Electronic Cash System (2008). https://bitcoin.org/bitcoin.pdf. Accessed 30 Aug 2020

Sazb, N.: The Idea of Smart Contracts (1997). https://nakamotoinstitute.org/the-idea-of-smart-contracts/. Accessed 30 Aug 2020

Visconti, R.M.: Blockchain Valuation: Internet of Value, digital networks and smart transactions (2019). https://www.researchgate.net/publication/329916782. Accessed 30 Aug 2020

Using Education Technologies to Accommodate Vocational and Professional Education Training (VPET) Students' Attributes

Kwan-Keung Ng[1](✉), Ricky Yuk-Kwan Ng[2], and Rechell Yee-Shun Lam[2]

[1] University of Sunderland, Sunderland, UK
steven.ng@sunderland.edu.hk
[2] Vocation Training Council, Hong Kong, Hong Kong SAR, China
{rickyng,rechelllam}@vtc.edu.hk

Abstract. Hands-on practices in vocational and professional education and training's (VPET) workshops and workplaces are the major learning and teaching activities. Interestingly, studies found that VPET's students prefer online learning and e-resources to complement and supplement their study. VPET students also found technology-enhanced learning (TEL) motivates learning and able to provide a timely review at anytime and anywhere. This paper shares the TEL strategies and best practices of one of Hong Kong's largest VPET institutions. Findings indicated the possibility and needs of using TEL to accommodate VPET students' attributes. Responses from the survey revealed that VAR best fitted psychomotor skills oriented students while also benefited cognitive oriented students to develop analytical skills. The qualitative feedback further revealed that the VAR assisted courses were appropriate and effective to motivate students, enhance knowledge, promote problem-solving and practical skills at their own learning pace. Most significantly, this study highlighted VPET students' attributes, learning styles and preferences as salient points for future studies.

Keywords: Vocational and professional education and training (VPET) · Students' attributes · Blended-learning · Technology-enhanced learning (TEL) · Gamification · Augmented and virtual reality (AVR)

1 Introduction

Studies revealed that using technology for learning and teaching improves learning effectiveness, advances knowledge and promotes student-led learning [1, 5, 6, 8, 11, 13, 14, 26, 29, 31–33]. Salmon [27] suggested that e-learning as an approach and pedagogy should now be incorporated into the education domain as a supportive mechanism for continuing development. For the past decade, the profusion of online learning platform and e-resources changed the concept of acquisition of knowledge and constituted a paradigm shift in learning and teaching. Knowledge on the internet appears in various forms of technology-enabled media such as online lectures, discussion forums, instruction videos and various means of interactive technology-enhanced learning (TEL). Arguably, the

L.-K. Lee et al. (Eds.): ICTE 2020, CCIS 1302, pp. 349–361, 2020.
https://doi.org/10.1007/978-981-33-4594-2_29

theory and lecture-based teaching in universities as well as their students' broadminded attributes and cognitive abilities in academic studies are the enabling factors for the well acceptance and suitability of online learning. Vocational and professional education and training (VPET) emphasises the mastery of hands-on dexterities and functional skills as the major learning and teaching activities in VPET's institutions, the conversion of hands-on practices into online learning delivery mode and the development of online resources need extra efforts to ensure the effectiveness on learning. Given the above, what are the e-learning strategies and educational technologies that will accommodate VPET's students' attributes, learning preferences and habits as well to enable them to practice hands-on skills in the VPET learning environment? To what extent would the above enhance their knowledge, practical skills as well as to enrich their learning experience?

2 Students' Attributes and Learning Preferences

Bloom's taxonomy originated in 1956 [3] classified educational learning objectives into three hierarchical models as cognitive (knowledge based), affective (emotion-based) and psychomotor (action-based) domains. Most teachers well accepted that that university students' broadminded attributes and cognitive abilities contribute to their excellence in academic studies; while vocational and professional education and training's (VPET) students are less excel in academic studies but merit in hands-on dexterities and functional skills. Their diligent attributes and psychomotor abilities shaped their learning preferences and habits. Whilst Bloom's taxonomy still valid after all these years, with the advancement of technologies and the arrival of digital culture that changed the ways people teach and learn as well as living, entertaining and consumption habits, the learning styles of the 'Post 90s' or 'Generation Z' can be summarised as follows:

- Perform multi-task, prefer visual image communications, auditory discussions and kinaesthetic hands-on experience over textual reading
- Be timely connected in any forms of electronic gadgets and social networks for instant exchanges of messages
- Collaborate, share knowledge and learn effectively through online social interactions

The increasing inclination to online or web-based visuals information to a certain extent related to the students' learning styles. The VARK learning style (visual, aural, read/write and kinaesthetic) by Fleming [10] provides well analysis of the learning styles and strategies to accommodate different students' needs. Fleming found that 55–65% of the populations are multimodal students possess with more than one learning styles. To a certain extent, art and design students are mostly visual-oriented; humanity and social science students are read/write oriented while the majority of technical and vocational students are kinaesthetic oriented. Interestingly, in a study Ng and Lam [20] found that VPET teachers remained skeptical about the suitability and effectiveness of e-Learning and other forms of TEL. Their perception deeply rooted in VPET's distinctive hands-on practices, workshop drillings and on-the-job learning. VPET teachers asserted that e-learning could easily well-adopted in universities because of their theory and knowledge-based teaching. Although studies by researchers in different academic disciplines [1, 5, 6, 8, 11, 13, 14, 26, 29, 31–33] indicated that online learning benefits students' learning in a range of areas; Hau [12] contended that most of the academic advanced students preferred paper-based and book-based learning rather than e-learning. Hau further asserted that the online learning effectiveness relied on the learning contents, modes and forms of delivery as well as pedagogies. Nevertheless, the above indicated that learning attributes, learning styles and learning preferences are closely related and technologies seem to be the promising enablers to enhance and enrich learning and teaching experiences to accommodate different learning needs and attributes.

3 Innovative Pedagogies and Technology Enhanced Learning (TEL)

Schuck, Kearney and Burden [28] asserted that mobile technology created a 'third space' for learning in addition to the traditional 'first space' learning in school such as lectures, tutorials, laboratory tests and the 'second space' out school field trips, site visits learning. The third space that Schuck, Kearney & Burden proclaimed is online learning in different spaces and locations, for example, in transportations during commuting, in café during coffee breaks or just-in-time learning before classes, examinations or meetings etc. Ng [19] further extended the above concept into a 'fourth space' enabled by technologies to promote workplace learning, such as practicum, on-the-job training or as named in Germany the 'dual system' to facilitate better blended-learning in VPET. VPET students may attend online lectures, video conferences, tutorials or sharing of practices in different workplaces other than school or home. Paired with online technologies, the flipped classroom approach [2] opens up a variety of innovative pedagogies and interactive learning activities, could it be online/offline videos, quizzes, assessments, discussion forums, micro classes and massive open online courses (MOOC) on learning management platforms or other more advanced learning resources such as virtual/augmented reality (VAR), three dimensional gamification, quick time virtual tour or google earth positioning system, just to name a few. TEL provides self-paced, self-led, just-in-time learning as well as repeated viewing to consolidate knowledge, enrich and enhance learning effectiveness.

A government funded TEL project and survey led by the three authors of this paper revealed that successful TEL resources bear the factors of interesting contents, good instructional design, up-to-dated technologies, buy-in from teachers and students and most importantly, well suit the needs and appropriateness for learning. The survey accompanied with the project aimed to examine learning needs and teaching practices. 343 students and 20 teachers from 3 VPET institutions participated in the survey to draw views on introducing innovative pedagogical practices involving technology-mediated forms of delivery. The questionnaire items for the 343 students addressed learning practices, learning needs and preferences as well as learning support. The questionnaire items for the 20 teachers related to teaching practices, teaching needs, preferences and teaching support. As the questionnaires were paper-based and the surveys were conducted in classes, the return rate is 100%. Findings of the survey revealed that TEL would be the appropriate innovative pedagogical practices for VPET. Using media and representations (text, graphs, tables, audio, videos, animations and interactive dynamic visuals) would accommodate learning and teaching needs for students with different learning styles. It also showed that new technologies enabled blended-learning and further enhance students' motivation and interaction. In addition, the survey also suggested that flexibility; guidance, collaboration and training would be able to accommodate VPET students, teachers and workplace mentors' learning and teaching needs [18, 21, 22]. Although the results shed directions and strategies on pedagogical practices for TEL and e-learning, the application of them in VPET is still in question. Furthermore, how would and possibly TEL accommodate VPET students' attributes for effective learning?

4 E-Learning Environment in Vocational and Professional Education and Training (VPET)

The process of work in real workplaces together with VPET teachers' real work hands-on process are the essences of VPET [9]. Ng and Lam [20] found out that most of the VPET training institutions' teaching staff were reluctant to TEL because a majority of VPET's trainings were hands-on and conducted in workshops with specific facilities and settings. VPET teachers also skeptical about the effectiveness of online learning particularly on students' efforts on self-regulated learning. Although Ng, Ng and Liu [23] claimed that flexible and self-learning suits workplace learning and training well because it enables flexible and timely learning without physical boundaries; they also realised that technologies are constantly under rapid changes, development of TEL learning resources must not solely rely on technologies but more attention shall be given on pedagogies and instructional design.

5 Case Study

One of the authors of this paper led an institutional wide 2-year project in 2018 and 2019 to promote the implementation of TEL in VPET. The project deliverables included mini-MOOCs for various trades/disciplines and a number of VAR assisted courses. With an aim to enhance and enrich VPET students' learning, TEL in VPET focuses on the mastery of professional subject knowledge, practical and soft skills. With rich contents, personalisation of learning, timely communications, interactive activities and all-time accessibilities via multi-sensory e-learning resources; TEL aims to accommodate VPET students' attributes and learning styles. To develop and implement the e-learning resources for effective learning and teaching, a blended-learning approach [25], integrating a combination of traditional learning with web-based learning activities was adopted. The TEL activities also articulated Bloom's digital taxonomy as devised by Churches [7]. Online videos, documents and self-study resources were used to articulate 'Remembering'; online quizzes and assessment to address 'Understanding'; online discussions and forums to promote 'Analysing'; online rating votes and appraisals for 'Evaluating' while online collaborations and projects accelerated 'Creating'. After discreet discussions and planning, the rule of thumb of converting the existing courses into e-learning resources was 'same course, same content with different delivery'. Using blend-learning and flipped classroom approach, the face-to-face time was re-structured by incorporated more active learning, communication, collaboration and engagement activities. Pedagogically, TEL (VAR, gamification, live broadcasting etc.) and mobile devices and tools such as smartphones, iPads, tablets and a range of EduTech Apps were used to motivate and enhance engagement.

Lessons were divided into face-to-face (in-class) and out-of-class (pre/post-class) activities. Applying the blended-learning approach and concept of flipped classroom enabled by technologies, students are recommended to acquire prior knowledge from online e-resources and finish the required readings, videos, tests, quizzes etc. uploaded in the learning management platform as pre-class activities. During face-to-face in-class sessions, students were engaged in collaborative group work for feedback from teacher and peers and master their hands-on practices using TEL tools such as classroom response system, VAR on their devices to apply and consolidate their prior learnt knowledge. Post-class extended learning such as revision, assessment on learning and management platform or real-time discussions using social media were adopted to promote situated learning and community of practice [15, 16] to build mutual trust, to construct and share knowledge for reflective learning. The instruction design of the lessons aligned Bloom's digital taxonomy where online self-study resources were used for 'Remembering'; online quizzes and assessment to address 'Understanding'; in-class discussions, practices to promote 'Analysing' and 'Applying'; and in-class and online participation, reflection and sharing were for 'Evaluating' and 'Creating'. The e-resources and activities also tied to Flemming's VARK to accommodate VPET students' attributes and learning styles. Converting traditional passive teaching into active blended-learning and embedding TEL in the activities would able to motivate learning intentions and facilitate active peer learning to suit the learning styles and needs of VPET students.

6 Findings and Discussions

Three VAR assisted courses were selected to collect feedback because when compared with the video-based mini-MOOC lectures, VAR's kinaesthetic and multi-sensory nature were more well-aligned with VPET students' attributes and learning preferences. Surveys were conducted and questionnaires were sent to students after the piloting of the three VAR assisted courses (Crime Scene Inspection (n = 25), 360 VR for Tour Guide Training (n = 17) and Safety Awareness in Construction Workplaces (n = 18). The survey results are shown in the following tables (Tables 1, 2 and 3).

Findings showed that in general, majority of the students agreed on the smooth accessibility, user-friendly layout and interface, well-designed learning content and successful outcomes after using the VR application for learning. With a consistency of rating of 80% and above on the 'agree', students were pleased with the pilot-run classes of Crime Scene Inspection and Safety Awareness in Construction Workplaces. Interestingly, the lower percentages (from 40% and above) on 'agree' indicated that students are less satisfied with the 360 VR for Tour Guide Training. As indicated by the results, when comparing with the ratings of Crime Scene Inspection and Safety Awareness in Construction Workplaces, students of Tour Guide Training were not keen on the VR environment and training. The reason for the above might resulted from the nature of the trade subjects and the accessibility of the sites. There was a mutual understanding

Table 1. Crime scene inspection

	Questions	Disagree	Agree
Total number of students: 25 Total number of responses: 25 Survey return rate:100%			
Accessibility			
1	I was able to gain authentic workplace experience in the VR environment (e.g. simulated tour) effectively	12%	88%
2	I was able to access the learning materials on the VR application (e.g. demonstration videos & graphics) easily	8%	92%
Layout and Interface			
3	The navigation of the VR interface was user-friendly	16%	84%
Learning Content			
4	The learning objectives were clearly stated in each VR activity (e.g. tour guiding procedures)	8%	92%
5	The learning materials and contents in VR-assisted lesson were well structured	8%	92%
6	The VR learning contents were effectively presented and explained clearly	8%	92%
7	The numbers of VR activities were enough	4%	96%
8	The VR learning activity was best for demonstrating practical skills	12%	88%
9	The VR learning activity was best for visualizing the concepts and theories	4%	96%
10	The VR learning activity was best for simulating the workplace environment	8%	92%
11	The VR activities presented the subject in an interesting way	8%	92%
12	The duration of the VR activities was suitable	12%	88%
Outcomes			
13	My overall knowledge of this subject was enriched after engaged in the VR-assisted Learning	16%	84%
14	The VR-assisted Learning motivated me to learn	16%	84%
15	I was more engaged in learning with the VR-assisted Learning than the face-to-face classroom teaching.	8%	92%
16	The VR-assisted Learning aroused my learning interest	8%	92%
17	Overall, I had a good learning experience with the VR-assisted Learning	8%	92%

amongst students and teacher that crime scenes and construction sites were prohibited to the public and had limited access to external parties unless permissions were granted. The tourist sites in nature were accessible for public and students still wanted to experience the actual places in person so that they could interact with peers, teachers and other tourists for hands-on-practices. A noteworthy point that the responses from a number of questionnaire items in Safety Awareness in Construction Workplaces' survey; item 2. I was able to access the learning materials on the VR application (e.g. demonstration videos & graphics) easily; item 4. The learning objectives were clearly stated in each VR activity (e.g. tour guiding procedures); item 7. The numbers of VR activities were enough; item 8. The VR learning activity was best for demonstrating practical skills; item 10. The VR learning activity was best for simulating the workplace environment; item 12. The duration of the VR activities was suitable and item 17. Overall, I had a good learning experience with the VR -assisted Learning were rated 100% 'agree'. Out of the three courses, the Safety Awareness in Construction Workplaces was an engineering course and the trade skills were technical driven and students' attributes were kinaesthetic oriented and fell into Bloom's psychomotor (action-based) domain, while the other two courses in this study were more into humanistic and analytical thinking (read/write) as in Bloom's cognitive (knowledge based) domain (Table 4).

Qualitative feedback revealed that the VAR assisted courses were appropriate and effective to motivate students, enhance knowledge, promote problem-solving and practical skills at their own learning pace. Feedback also reflected a certain extent of learning

Table 2. 360 VR for tour guide training

Total number of students:17 Total number of responses: 17 Survey return rate:100%			
	Questions	Disagree	Agree
Accessibility			
1	I was able to gain authentic workplace experience in the VR environment (e.g. simulated tour) effectively	24%	76%
2	I was able to access the learning materials on the VR application (e.g. demonstration videos & graphics) easily	24%	76%
Layout and Interface			
3	The navigation of the VR interface was user-friendly	24%	76%
Learning Content			
4	The learning objectives were clearly stated in each VR activity (e.g. tour guiding procedures)	24%	76%
5	The learning materials and contents in VR-assisted lesson were well structured	24%	76%
6	The VR learning contents were effectively presented and explained clearly	18%	82%
7	The numbers of VR activities were enough	47%	53%
8	The VR learning activity was best for demonstrating practical skills	53%	47%
9	The VR learning activity was best for visualizing the concepts and theories	47%	53%
10	The VR learning activity was best for simulating the workplace environment	41%	59%
11	The VR activities presented the subject in an interesting way	24%	76%
12	The duration of the VR activities was suitable	41%	59%
Outcomes			
13	My overall knowledge of this subject was enriched after engaged in the VR-assisted Learning	12%	88%
14	The VR-assisted Learning motivated me to learn	47%	53%

(*continued*)

Table 2. (*continued*)

Total number of students:17 Total number of responses: 17 Survey return rate:100%			
	Questions	Disagree	Agree
15	I was more engaged in learning with the VR-assisted Learning than the face-to –face classroom teaching	47%	53%
16	The VR-assisted Learning aroused my learning interest	35%	65%
17	Overall, I had a good learning experience with the VR-assisted Learning	35%	65%

efficiency. Students appreciated the VAR experiences enabled them a deeper understanding of the locations, applied knowledge in the virtual environment and the revisited of information and resources at home. Most students also found the user interfaces convenient, the applications easy to use and control. They were delighted by the "authentic videos and vivid cartoon" as well as "vocal narration and text navigation". Further comments revealed that better gadgets design would help. On lessons and activities arrangements, students found clear instructions and guidance with interesting exercises and activities. They also agreed that VAR activities added more interactivity in regular learning and teaching. Some students asserted that "VAR learning cannot substitute face-to-face teaching" and further mentioned that blend-learning would be adopted on a proportion of 2/3 classroom teaching and 1/3 VAR learning. Apparently, blended VAR courses with kinaesthetic and multi-sensory learning activities seemed most appropriate to cater to the learning needs and attributes of VPET's students. It also echoes the previous studies that Tang, J., Pang, W., Wong, P. et al. [30] and Lee, C., Lam, S., Liu, A. et al. [17] had conducted on using flexible learning in nursing education to enhancement psychomotor skills. Interestingly, this study also found that VAR best fitted the learning in the engineering course which also involved with a lot of psychomotor skills. Nevertheless, VAR also benefited VPET students to improve their cognitive skills as showed in the findings of the other two VPET courses in this study. In sum, the blended-learning and flipped classroom approaches not only motivated and raised learning interests, enhanced subject knowledge and enriched learning experiences but also provided all-time accessibilities for self-paced learning. The instruction design for the learning and teaching activities in the VAR assisted courses in this study also articulated Bloom's digital taxonomy to facilitate 'Remembering', 'Understanding', 'Analysing', 'Applying', 'Evaluating' and 'Creating'.

Table 3. Safety awareness in construction workplaces

Total number of students: 22 Total number of responses: 18 Survey return rate: 82%			
	Questions	Disagree	Agree
Accessibility			
1	I was able to gain authentic workplace experience in the VR environment (e.g. simulated tour) effectively	6%	94%
2	I was able to access the learning materials on the VR application (e.g. demonstration videos & graphics) easily	0%	100%
Layout and Interface			
3	The navigation of the VR interface was user-friendly	6%	94%
Learning Content			
4	The learning objectives were clearly stated in each VR activity (e.g. tour guiding procedures)	0%	100%
5	The learning materials and contents in VR-assisted lesson were well structured	6%	94%
6	The VR learning contents were effectively presented and explained clearly	6%	94%
7	The numbers of VR activities were enough	0%	100%
8	The VR learning activity was best for demonstrating practical skills	0%	100%
9	The VR learning activity was best for visualizing the concepts and theories	6%	94%
10	The VR learning activity was best for simulating the workplace environment	0%	100%
11	The VR activities presented the subject in an interesting way	6%	94%
12	The duration of the VR activities was suitable	0%	100%
Outcomes			
13	My overall knowledge of this subject was enriched after engaged in the VR-assisted Learning	6%	94%
14	The VR-assisted Learning motivated me to learn	17%	83%

(*continued*)

Table 3. (*continued*)

Total number of students: 22 Total number of responses: 18 Survey return rate: 82%			
	Questions	Disagree	Agree
15	I was more engaged in learning with the VR-assisted Learning than the face-to –face classroom teaching	11%	89%
16	The VR-assisted Learning aroused my learning interest	6%	94%
17	Overall, I had a good learning experience with the VR -assisted Learning	0%	100%

Table 4. Summary of qualitative feedback

1. *Appropriateness and effectiveness*
Raise my interest in learning, enhance my problem-solving capability, create an authentic learning environment to enrich my study and enhance workplace etiquette
VR application provided more on-site information that enhanced my understanding in the site
VR and AR content are more vivid than the theories taught in the face-to-face classroom
Can navigate in the VR environment by myself and have more learning autonomy
It is like going to a real crime scene
Enhanced my safety awareness by identifying good and bad practices
2. *Learning efficiency*
Abundant information embedded in the applications and I could revise the information at home
Appreciated the VR and AR experience for they provided an authentic environment to apply knowledge
Have a deeper understanding of the locations of the sightseeing spots
The use of applications was more handy and free from the weather and physical constraints (e.g. go into forbidden scenes or being blocked by the other tourists in the real location)
A clear illustration of the construction site components
3. *User interface*
VR applications are easy to control and convenient to use
AR application is easy to use as the whole building procedure is directly shown in the application
Good to have authentic videos and vivid cartoon in the AR application
Good to have vocal narration and text navigation
VR is more difficult to use as it needs to adjust the monitor size
A little bit difficult in controlling AR as the phone must be placed on the accurate spots to view the animation in the accurate place
4. *Lessons and activities arrangement*
Teacher's guidance is sufficient and instruction is clear
Interesting, in-class exercise/activities
Sufficient time to explore/learn with the VAR applications
VAR Learning cannot substitute face-to-face teaching (Students can ask questions directly and the visions were more clear and authentic when they visit the real locations)
Good to have some VAR activities to add more interactivity in regular learning and teaching
Suggest to have a mixed proportion of 2/3 classroom teaching and 1/3 VAR learning

7 Implications and Conclusion

The shift of students' characteristic, preferred learning styles and attributes provided a chance for teachers to explore and think about how to make use of technology to facilitate students' learning. Riding on the result and experience of this study, blended-learning and TEL assisted courses in the future may extend to a "personalised, learner-centred, situated, collaborative, ubiquitous and contextual mobile learning environment" [24, p.2] to deepen students' engagement and collaboration. The formation of 'learning community' and 'situated learning' will accelerate the concept of "amplification of learning" [29] enabled by the networked online learning environment. This type of "amplification of learning" [15, 16] would be able to "deepen levels of students' engagement and collaboration within the digital learning environment….to provide a coherent wholesome learning experience" [4, p.1]. To conclude, this study addressed the possibility of using TEL to accommodate VPET students' attributes. Findings of this study also shed lights on the importance of instructional design, innovative pedagogies, learning and teaching strategies and approaches to enhance knowledge, soft and practical skill as well as to enrich learning experiences. Most significantly, this study highlighted VPET students' attributes, learning styles and preferences as salient points for future studies.

References

1. Barnett, L.: Teacher off: computer technology, guidance and self-access. System **21**(3), 295–304 (1993)
2. Bergmann, J., Sams, A.: Flip Your Classroom: Reach Every Student in Every Class Every Day, pp. 120–190. International Society for Technology in Education, Washington DC (2012)
3. Bloom, B.S.: Reflections on the development and use of the taxonomy. In: Rehage, K.J., Anderson, L.W., Sosniak, L.A. (eds.). Bloom's Taxonomy: A Forty-Year Retrospective. Yearbook of the National Society for the Study of Education, vol. 93. National Society for the Study of Education, Chicago (1994)
4. Boulos, M.N.K, Maramba, I., Wheeler, S.: Wikis, blogs and podcasts: a new generation of web-based tools for virtual collaborative clinical practice and education. BMC Med. Educ. **6**, 41 (2006). http://www.biomedcentral.com/1472–6920/6/41. 14 June 2013
5. Chau, J., Cheng, G.: ePortfolio, Technology, and Reality: a Reality Check. J. Interact. Learn. Res. **21**(4), 221–237 (2010)
6. Chau, J., Cheng, G.: Exploring the relationship between students' self-regulated learning ability and their ePortfolio achievement. Internet High. Educ. **17**(2013), 9–15 (2013)
7. Churches, A.: Bloom's Digital Taxonomy (2008). https://www.researchgate.net/publication/228381038_Bloom's_Digital_Taxonomy/link/5d2ab7ff458515c11c2e24fb/download. Accessed 25 Sept 2020
8. Edbert, J.: Supporting Learning with Technology: Essentials of Classroom Practice. Pearson, Upper Saddle River (2009)
9. Diep. P.; Hartman, M.: Green Skills in Vocational Teacher Education-A Model of Pedagogical Competence for a World of Sustainable Development (2016). www.tvet-online.asia. Accessed 10 Jan 2019
10. Fleming, N.: Teaching and learning styles: VARK strategies. IGI Global (2001)
11. Foreman, J.: Next generation: educational technology versus the lecture. Educause Rev. **38**, 12–22 (2003)

12. Hau, K.T.: Policy Café. Our Hong Kong Foundation (2020). https://www.facebook.com/1183623165011878/posts/3761550680552434?sfns=mo. Accessed 16 March 2020
13. Jenings, S.E., Weatherly, M.G.: Social network in education and connectivism learning. J. Res. Bus. Inf. Syst. **6**, 50–72 (2013)
14. Keats, D., Schmidt, J.P.: The genesis and emergence of education 3.0 in Higher Education and Its Potential for Africa (2007). http://firstmonday.org/ojs/index.php/fm/article/view/1625/1540. Accessed 22 June 2013
15. Lave, J.: Situating learning in communities of practice. Persp. Soc. Shared Cogn. **2**, 63–82 (1991)
16. Lave, J., Wenger, E.: Situated Learning: Legitimate Peripheral Participation, pp. 33, 29, 40. Cambridge University Press, Cambridge (1991)
17. Lee, C., Lam, S., Liu, A., et al.: Enhancement of psychomotor skills through flexible learning in nursing education.: a quasi-experimental study on respirator donning. In: Yuen, K.S., Lee, K.C. (eds.) Emerging Modes and Approaches in Open and Flexible Education. Open University of Hong Kong Press, Hong Kong (2014)
18. Ng, R.: A report on the cross-institutional study of vocational education and training (VET) students' learning needs as well as teachers and workplace mentors' teaching practices" for the quality enhancement support scheme (QESS) Project: development of effective pedagogical practices and a cross-institutional online sharing platform for Hong Kong's VET. https://vpetcity.vtc.edu.hk/doc/8_study_report.pdf. Accessed 30 Oct 2018
19. Ng, R.: Omniscience learning space: a new dimension of learning. In: VPET Conference Series in the Learning and Teaching Expo 2017, Hong Kong Convention Centre, 13th to 15th December, 2017 (2017)
20. Ng, R., Lam, R.: Student learning preferences for and perceptions of blended learning: a case study in MOOCs for vocational education in Hong Kong. In; Hong Kong International Conference on Education, Psychology and Society (HKICEPS), Eaton Hotel, Hong Kong, 14–16 December 2016 (2016)
21. Ng, R., Lam. R., Ng, K., Lai, K.: A cross-institutional study of vocational and professional education and training (TVET) students and teachers' needs of innovative pedagogical practices. In: Proceedings of 2016 International Symposium on Educational Technology (ISET), Beijing, pp. 101–105 (2016)
22. Ng, R., Lam, R., Ng, K., Lai, K.: A study of vocational and professional education and training (TVET) students and teachers' preferred support for technology based blended learning. In: Proceedings of the International Symposium on Educational Technology, Hong Kong, 27–29 June 2017 (2017)
23. Ng, R.Y., Ng, K.K., Liu, B.: An empirical study on the usefulness, effectiveness and practicability of vocational and professional education and training's (VPET) open educational resources (OER) in the hotel industry. In: Cheung, S.K.S., Lee, L.-K., Simonova, I., Kozel, T., Kwok, L.-F. (eds.) ICBL 2019. LNCS, vol. 11546, pp. 265–276. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-21562-0_22
24. Oller, R.: The Future of Mobile Learning. EDUCAUSE. Centre for Applied Research (2012)
25. Oliver, M., Trigwell, K.: Can 'blended learning' be redeemed? E-learning and Digit. Media **2**(1), 17–26 (2005)
26. Richardson, W.: Blogs, Wiki, Podcasts, and Other Powerful Web Tools for Classroom, 2nd edn. Corwin Press, Thousand Oaks (2009)
27. Salmon, G.: Flying not flapping: a strategic framework for e-learning and pedagogical innovation in higher education institutions. ALT-J (now Research in Learning Technology) **13**(3), 201–218 (2005)
28. Schuck, S., Kearney, M., Burden, K.: Exploring mobile learning in the third space. Technol. Pedagogy Educ. **26**(2), 121–137 (2017)

29. Siemens, G.: Connectivism: a Learning Theory for the Digital Age (2004). http://www.elearnspace.org/Articles/connectivism..htm. Accessed 22 June 2013
30. Tang, J., Pang, W., Wong, P., et al.: Enabling a flexible university education through an augmented reality interactive interface. In: Yuen, K.S., Lee, K.C. (eds.) Emerging Modes and Approaches in Open and Flexible Education. Open University of Hong Kong Press, Hong Kong (2014)
31. Wignall, E.: Media, minds and methods: linking online instruction and media for maximum effect. Paper Presented at the 21st Annual Conference on Distance Teaching and Learning, May 2006 (2006). http://www.uwex.edu/disted/conference/Resource_lobrary/proceedings/05-1898.pdf. Accessed 22 June 2013
32. Williams, P.J.: Design: the only methodology of technology? J. Technol. Educ. **11**(2), 48–60 (2000)
33. Zemsky, R., Massy, W.F.: Thwarted innovation-what happened to e-learning and why? a final report for the weatherstation project of the learning alliance at the University of Pennsylvania in cooperation with the Thompson Corporation, pp. 1–76. The University of Pennsylvania, Pennsylvania. http://www.educause.edu/library/resources/thwarted-innovation-what-happened-e-learning-and-why. Accessed on 14 June 2013

Recognition of Roles of Variables Based on Deep Learning Technologies

Andrew Kwok-Fai Lui(✉), Tsz-Tik Lui, Tommy Cheuk-Hin Leung, and Ho-Yin Chan

Department of Technology, The Open University of Hong Kong, Ho Man Tin, Kowloon, Hong Kong SAR, China
alui@ouhk.edu.hk, {g1212688,g1186142,g1175028}@study.ouhk.edu.hk

Abstract. Using variables in regulating computing procedures is a key competence of structured programming. This tacit knowledge, which is highly repetitive and transferrable between programming tasks, may be succinctly described as the roles of variables. The role of a variable defines the pattern of the uses and changes of variable for computing procedures. This notion is an effective pedagogical tool for providing feedback on how to fix faulty computing procedures linked to misuse of variables. This paper proposes the automation of variable role recognition and describes an evaluation study for a variable role classifier based on deep neural network architectures. The classifier is designed to learn the variable roles from execution-time traces of variables. Experimental results showed the classification accuracy reached over 0.95. Other findings indicated that the variable roles revealed diverse temporal characteristics in their execution-time traces, and specifically designed deep learning architectures would be needed for some individual roles. Automated variable role recognition will find applications in intelligent tutoring systems and the feedback to programs will significantly improve learning effectiveness.

Keywords: Programming · Novice programmers · Roles of variables · Machine learning · Deep learning

1 Introduction

Computer programming is the task of designing an executable program that uses, processes and generates data for specific computational objectives. Mastery of computer programming embroils the competency in a number of programming skills. One of them is the use of variables in regulating computing procedures. Variables are basically labelled data, and the standard operations on variables include initialization, updating and referencing. A functional computing procedure involves thoughtful use of variables. There are a number of use cases that are found commonly occurring in different computing procedures, and variables are playing a specialized role in each of these use cases. The role of variable describes a characteristic pattern of variable operations that can be observed from the changing referenced data (Sajaniemi 2002). For examples, a variable playing the *stepper* role is critical for stopping a repeating procedure and it assumes a

L.-K. Lee et al. (Eds.): ICTE 2020, CCIS 1302, pp. 362–374, 2020.
https://doi.org/10.1007/978-981-33-4594-2_30

sequence of numbers until the stopping condition is reached, and a *most-wanted holder* variable is needed for finding the largest value in the same procedure that traverses an array.

The notion of roles of variable is significant in the development of programming skills (Sajaniemi and Kuittinen 2005). A programmer familiar with the common use cases of variables can write programs more instinctively. The ability in recognizing the associated roles of variables for a computing procedure indicates competence. Expert programmers can recall known usage patterns, which is more efficient and reliable, than the cognitive effort in developing a new solution. Shi et al. (2018) found that in a tertiary programming course, learning outcomes in basic C programming knowledge, program comprehension, and program construction showed a significant improvement.

An error-free executable program would include variables playing all the roles that it should have assigned for its computing procedures. Programming errors can occur to anybody and especially novice programmers. Screening and recognizing the roles of variables in a program can help deduce if errors are due to omission or misplace of one or more variable roles. Variable roles are not explicitly labeled in the source program code and an understanding of the execution-time actions is needed for their recognition. Experienced programmers can usually recognize the roles of variables (Sajaniemi and Navarro Prieto 2005), but the same is not true for novices who may need help for their recognition.

1.1 The Aim and Structure of the Paper

This paper describes a study on automated recognition of the roles of variables. A variable role classifier based on a model of the execution-time characteristics of variables is to be presented. The data-driven deep learning approach has been adopted to learn the model from a sizeable set of execution-time traces of variables. The data-driven approach is effective for discovering the implicit knowledge of variable roles.

An example application for the technique is generation of feedback messages for programming exercises. The list of variable roles recognized in the model program answer can be used as the canonical reference for the assessment of student work. In case of programming errors due to misuse of variables, the feedback highlighting the omitted or misplaced roles will be handy for making corrections. The discovery of redundant variables will also be useful for improvement. The feedback may be provided during or after programming exercises, and the former is a form of intelligent and synchronous tutor. Online learning platforms are becoming popular for studying computer programming. Automated recognition of variable roles can become a key enabling technique for around-the-clock, scalable, cost-effective automatized instruction.

The remainder of the paper is structured as follows. Section 2 gives detailed background of the notion of roles of variables and discusses the technical issues and relevant techniques for their recognition. Section 3 describes the deep neural networks that are proposed for learning the execution-time models of variable roles. The experiment setup, including the preparation of training and testing data, will also be outlined. The next section describes the findings with regards to the training and the performance of the models. The last section discusses the significance of the findings and proposes directions for further research.

2 Background

This section analyses the problem of recognition of variable roles and makes clear the relevant technical issues that need to be resolved. The relevant technologies and research work are discussed, with the objective to justify the adoption of the data-driven deep learning approach.

2.1 Roles of Variables

In computer programming, variables are involved in defining data operations in a computing procedure. The variables are labeled data that play a specific role for the procedure. The role of a piece of data is the same role as the corresponding variable. Computer programming textbooks have long used names such as *counters* and *flags* to typify variable roles. Sajaniemi (2002) was the first attempt to formalize the notion and provided definitions for a set of common roles relevant to fundamental programming. Recently, Hosanee and Rana (2018) refined teaching with roles of variables using visualization, and Shi et al. (2018) integrated the notion into an integrated development environment with success. Demyanova (2013) applied the notion in software analysis. Some examples of the common roles can be found in Table 1.

Table 1. Examples of variable roles and their descriptions extracted from Sajaniemi (2002).

Roles	Descriptions
Constant	A variable that does not change after initialized with a value
Stepper	A variable successively takeing up a predictable sequence of values
One-way Flag	A variable that switches from an initial value to another value and remains unchanged
Gatherer	A variable that accumulates the effect of some values
Most-recent Holder	A variable holding the latest value in a sequence of relevant values
Most-wanted Holder	A variable that holds the best or most desired value among a sequence of relevant values

The descriptions for these roles are largely about value updates along the temporal dimension. The way how variables are used is irrelevant (Sajaniemi and Navarro Prieto 2005). Figure 1 illustrates the roles of the variables for a procedure that finds the largest value in an array. With the exception of the array variable that does not belong to any of the common roles, the three roles, namely *Stepper*, *Most-wanted holder* and *Constant*, are critical for this computing procedure. The example uses the C programming language, but the roles are also relevant in structured and object-oriented languages.

The temporal characteristics of the variables with respect to their roles can be illustrated through tracing their value changes at execution-time. A simple tracing model, called single trace, samples its values when a variable is updated. Figure 2 shows the single traces of two stepper variables with a distinctive increasing trend. The length of

```
int array[] = {5, 3, 6, 7, 2, 1, 5, 4};
int index = 0;
int highest;
int len = 8;

while (index < len) {
  if (index == 0 || array[index] > highest)
    highest = array[index]; /* A */
  index += 1; /* B */
}
```

Variable Names	Roles
array	---
index	Stepper
highest	Most-wanted Holder
len	Fixed Value

Fig. 1. Variables and their roles in a computing procedure that finds the largest value in an array of fixed length. Statements A and B contains variable update operations.

the tracing can be different because each time-step represents an update operation of a variable. Note that the two variables appear to have the same temporal characteristic, but their roles are actually different. The single traces may have limited differentiation capacity.

Single Traces

Time →

index	0	1	2	3	4	5	6	7	8

Time →

highest	?	5	6	7

Set Trace

Time →

array[index]	5	5	3	6	6	7	7	2	1	5	4	?
index	0	0	**1**	**2**	2	**3**	3	**4**	**5**	**6**	**7**	**8**
highest	?	**5**	5	5	**6**	6	**7**	7	7	7	7	7
len	8	8	8	8	8	8	8	8	8	8	8	8

Fig. 2. Two types of traces of variable values shown are sampled at the time of variable updates. The single trace (top) is a sampling of the variable `index` when it changes value in the program in Fig. 1. The single trace (middle) is for the variable `highest`. The set trace (bottom) is a sampling of all 4 variables when any one of them was updated. The shadow boxes indicate the updated variables for that time-step.

The context, such as the values of other variables, is significant for defining the role of a variable. Two or more variables may be working collaboratively in a computing procedure. Characteristic patterns may only be revealed if the variables are examined together. Another tracing model, called the set trace, samples the values of the whole set of variables when one of them is updated. Figure 2 also shows the set trace of the four variables in the program in Fig. 1. This tracing allows the revelation of a pattern involving three variables, that `highest` obtained the value from `array[index]` in between the updates of `index`. The set trace contains significantly richer information than variable trace with respect to the recognition of variable roles.

The notion is concerned with deep structure of programs and yet experienced programmers can recognize these roles (Sajaniemi and Navarro Prieto 2005). A study into the acceptability of the notion and the recognition of individual roles found that computer science educators could agree between them 90 percent of role assignment in sample programs (Ben-Ari and Sajaniemi 2004). The novice teachers also performed at a similar level. The subjects achieved this remarkable correlation level after just a short tutorial and training session. The roles are intuitive, and the associated tacit knowledge is shared by experienced programmers.

2.2 Recognition of the Roles of Variables

Computer recognition of the roles of variables is a specialized and less-studied research area in automated program comprehension. It is artificial intelligence that is capable of understanding computer programs like experienced programmers. In addition to teaching and learning of programming, automated program comprehension is also useful for recovery of system design, program repairs (Gupta et al. 2017), and program synthesis (Kant 2018). A critical process in automated program comprehension is the extraction of features that represent the knowledge of a program. There are three major ways to extract program features, including input-response analysis, syntax and structural analysis, and execution-time analysis.

Finding an effective set of features for role recognition is challenging. One approach is to exploit expertise of programmers. Bishop and Johnson (2005) identified 21 features of dynamic properties of variable and built a role recognizer called *RoleChecker* using hand-crafted rules constructed from the features. The features were extracted with a quasi-interpreter called the data flow analysis. A limitation of feature engineering is that feature set selection is still a matter of opinion. Gerdt and Sajaniemi (2006) used machine learning techniques to identify the required, tolerated, and non-tolerated features. The initial feature set may still omit some key features and potentially reduce the recognition accuracy.

Recently advances in deep learning technologies had revived interest in automated program comprehension. A deep learning neural network is capable of learning the features inherent in the data. This automated feature engineering can therefore circumvent the limitation of feature engineering. A variety of deep learning architectures, most of them based on variations and combinations of Recurrent Neural Networks (RNN) and Convolutional Neural Networks (CNN) have been proposed. The RNN and CNN are capable of building a spatial-temporal model of a sequential representation of the execution-time of programs. To the best knowledge of the authors, two architectures relevant to building dynamic models for variables were found in the literature. Devlin et al. (2017) uses a sequence-to-sequence architecture, which is based on multiple RNNs, to learn location patterns of variable use for program error localization. The most relevant work is Wang et al. (2017), that uses another multi-RNN based architecture to learn a spatial-temporal model of multiple variables from execution-time traces of variables.

2.3 Deep Learning Technologies for Variable Role Recognition

Large samples of publicly downloadable source code are available at large-scale repositories and online programming practice platforms. The abundance of data has made automated feature engineering an attractive option. The work by Wang et al. (2017) developed a program embedding, a spatial-temporal model of execution-time traces of variables, that was plugged into a downstream application of program repair. Such program embeddings are built for modularity and can be used for classification and other applications. The next section will delineate the use of deep neural networks for variable role recognition.

3 Method

The study on automated recognition of the roles of variables involved three major parts, namely data preparation, experimental platform setup, and deep learning network design. The premise of our recognition method is the pattern of variable and state traces in executable programs. An adequately large sample of such traces is needed for a deep network to learn the inherent features. A collection of fundamental level programs was used for the generation of the traces and the manual annotation of ground-truth variable roles.

The variable role recognition task was formulated as a classification problem. The number of classes equals to the number of variable roles to be recognized. A classifier is trained based on a model built to capture the features of variable and state traces of variables. A number of formulations for building the model were developed and evaluated. Computer programs for data preparation and the implementation of the temporal model formulations were used in the setup of an experimental platform. Figure 3 summaries the setup of the experimental platform and the design of the classifier for variable role recognition.

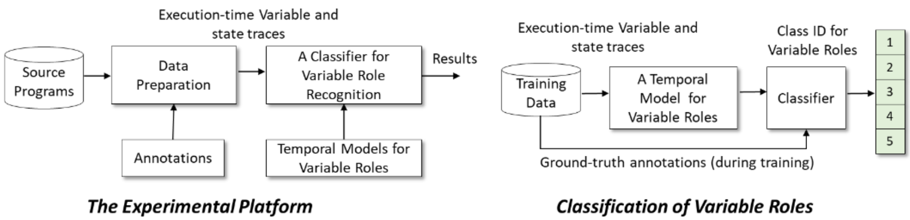


Fig. 3. The setup of the experimental platform used in the study is shown on the left. The procedure of the variable role classifier is illustrated on the right.

3.1 Formulations of Temporal Model for Variable Roles

Recall that traces of variables were presumed to contain features significant for variable role recognition. The purpose of the temporal model for variable role is to learn such features. A total of six formulations and models were evaluated. Each of the formulations represents a particular perspective of the temporal characteristics of variable traces and their interactions. Figure 4 describes the designs of the six deep learning models. The key features are discussed in the following paragraphs.

- *Classification.* A fully connected neural network (i.e. FC) that produces a probability distribution for the variable roles with a component called the *softmax* layer.

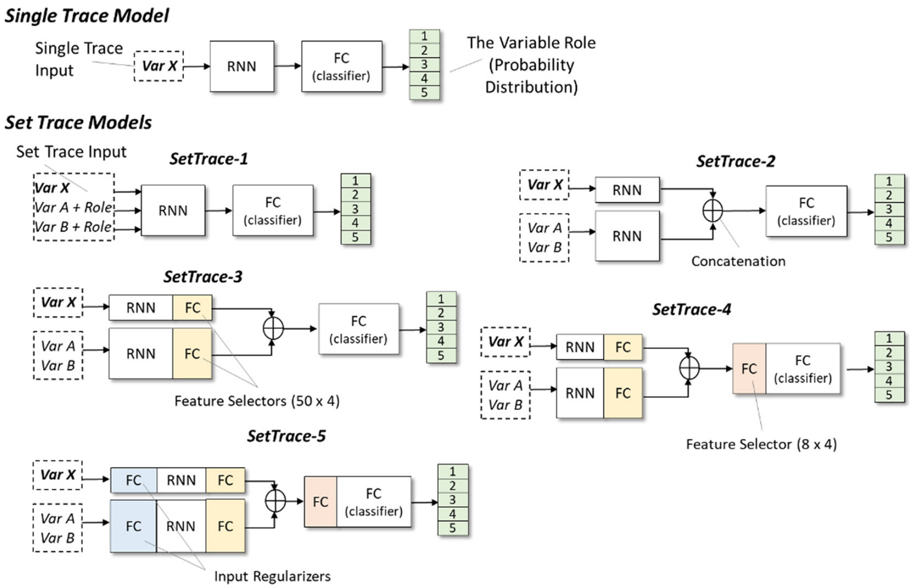


Fig. 4. The six formulations for feature learning associated with variable roles that were evaluated in this study. Var X is the variable annotated with a variable role. Var A and Var B represent the context variables for the annotated variable in the same program.

- *Single tracing and set tracing*. Each variable is considered as independent in a single trace which does not record any interaction between variables. Single traces of variables used as the input for the *Single Trace Model*. Set traces should better capture the interactions between variables that are significant for role recognition, and they are the input for *SetTrace*-1 to *SetTrace*-*5*.
- *Temporal modelling*. The traces have a temporal dimension. All models use a RNN which is a neural network type for learning temporal patterns.
- *Interaction between the annotated variable and the context*. The context, which includes the temporal patterns of other variables, is known to be significant for defining the role of a variable. *SetTrace-1* attempts to learn a combined model from state traces of the annotated variable and the context variables together. *SetTrace*-*2* to *SetTrace*-*5* attempt to learn a joint model of two sub-models, consisting of a temporal model of the annotated variable and a combined temporal model of the context variables.
- *Feature Selection*. The training of the classifier (i.e. the final FC components in the procedures) is easier if only the most relevant features are used. *SetTrace*-*3* uses two FC components to select features from the output of the RNN. *SetTrace*-*4* and *SetTrace*-*5* add another FC component to select features between the sub-models for the annotated variable and the context variables.
- *Input Regularization*. The input traces may include noises, such as the magnitude of values, which are irrelevant to role recognition. *SetTrace*-*5* uses FC components for the regularization and noise removal of each input parameter.

3.2 Preparation of Training and Testing Data

To prepare ground-truth annotated data for training the deep models and the classifier, a set of 110 C programs was selected from the same dataset used in Mou et al. (2016). The dataset came from a pedagogical programming open judge (OJ) system, which included 500 solution programs from 104 problems. The variable and state traces had to be extracting from executing those 110 C programs.

The extraction was facilitated by turning the C programs into C++ programs with two major changes. First, the variables to be traced were manually annotated with a user defined class called *Intercept* with which the role could be specified. A state trace could be made using another class called *StateIntercept*. Second, the assignment operators that update variables were overloaded so that their values could be logged at execution-time. The traces were written to JSON files and then packed into a Python NPZ file. An example of program annotation is shown in Fig. 5.

```
#include <../StateIntercept.h"
...

Intercept<int> ways(Intercept<int> a, Intercept<int> min) {
  if (a<min) return 0;
  Intercept<int> sum(0,3), i(-1,2);
  for (i=min; i<=sqrt(a*1.0); i++) {
    if (a%i == 0)
      sum = sum + ways(a/i, i);
  return sum+1;
}

int main() {
  Intercept<int> n(-1, 2), b;
  scanf("%d", &n);
  ...
}
```

Fig. 5. An annotated source program that had the variables turned into Intercept objects.

4 Results

Experiments were carried out to evaluate the deep learning approach to variable role recognition and to gain understanding of effective formulations of the deep learning architectures. The dataset consists of 8907 cases and it was split into a training set, a testing set, and a validation set in a 64%, 20%, and 16% respectively. The training set was used for training the deep learning models, the validation set was used for tuning the parameters during training and the testing set was used for evaluating the performance of model with unseen cases.

Table 2 shows the distribution of the variable roles that the dataset was annotated with. The *One-way Flag* and the *Constant* roles were combined into a class for the experiment because it was found at the early stage that these two roles could not be differentiated due to the method of data collection. This issue will be discussed later.

Table 2. Distribution of variable roles in the annotated dataset.

Roles	Stepper	Gatherer	One-way Flag & Constant	Most-Recent Holder
Frequency	2655	2502	2700	1050

4.1 Performance of the Single Trace Model

In this experiment, the *Stepper*, *Gatherer* and *One-way Flag* & *Constant* were extracted from the dataset for training and evaluating the model. Table 3 shows the performance of the *Single Trace* model using three metrics, precision, recall, and f1-score, for evaluating the classifier's quality. Due to the random nature of model training, the figures shown are the averaged results of three runs of the experiment. A score of 1.00 for the metrics indicate a perfect performance. The trained model was found to be very accurate in recognizing the three roles.

Table 3. Performance of the Single Trace Model.

Model	Training			Validation			Testing		
	Precision	Recall	F1 score	Precision	Recall	F1 score	Precision	Recall	F1 score
SingleTrace	0.964	0.964	0.964	0.953	0.953	0.953	0.959	0.959	0.959

The *Most-Recent Holder* cases were omitted in the experiment because the model was found incapable of differentiate this role from the rest. Set variable traces would provide the knowledge for the training of the deep models.

4.2 Performance of the Set Trace Models

In this experiment, the whole dataset was used for training and evaluating the five models with set traces as the input form. Table 4 shows the performance of the five Set Trace models. *SetTrace-2* was found to have the best performance with the training set, that is, seen cases. *SetTrace-5*, was found to perform best with the unseen testing set, which means it should have a better generalization capability.

Other major findings are listed below:

- Temporal modelling of the annotated variable should be separated from temporal modelling of the context variables. The temporal pattern of the annotated variable should be independently learned before its interaction with other variables as demonstrated by the superior performance of *SetTrace-2* to *SetTrace-5* over *SetTrace-1*, which attempted to build a combined set trace temporal model for all variables.
- Feature selection was found to be best placed after the concatenation of the separately trained temporal models for the annotated variable and the context variables as

Table 4. Performance comparison of the Set Trace Models.

Models	Training			Validation			Testing		
	Precision	Recall	F1 score	Precision	Recall	F1 score	Precision	Recall	F1 score
SetTrace-1	0.812	0.803	0.807	0.799	0.790	0.795	0.800	0.789	0.794
SetTrace-2	**0.984**	**0.965**	**0.975**	0.955	0.950	0.953	0.956	0.945	0.951
SetTrace-3	0.925	0.903	0.914	0.902	0.875	0.888	0.905	0.881	0.893
SetTrace-4	0.977	0.965	0.971	0.956	0.938	0.947	0.958	0.939	0.948
SetTrace-5	0.964	0.942	0.953	0.969	0.944	0.956	**0.971**	**0.949**	**0.960**

demonstrated by the improved performance of *SetTrace-4* and *SetTrace-5* over *SetTrace-3*. The feature selector has access to all the learned features from the annotated variable's trace and the context variable's traces at that stage.

- Input regularization was found to improve the prediction accuracy as shown in the improved performance of *SetTrace-5* over *SetTrace-4*. The regularization enabled *SetTrace-5* to better see through the variations and noises of the input variable traces and adapt to unseen input.

4.3 Recognition of the Variable Roles

Some variable roles are harder to be recognized than others. Figure 6 shows the confusion matrices for *SetTrace-1* and *SetTrace-5*. The recognition of the *One-way flag* role looked easier as even the relatively poor *SetTrace-1* did well. The accuracy of recognizing the *Gatherer* role was found to be lowest even for *SetTrace-5*.

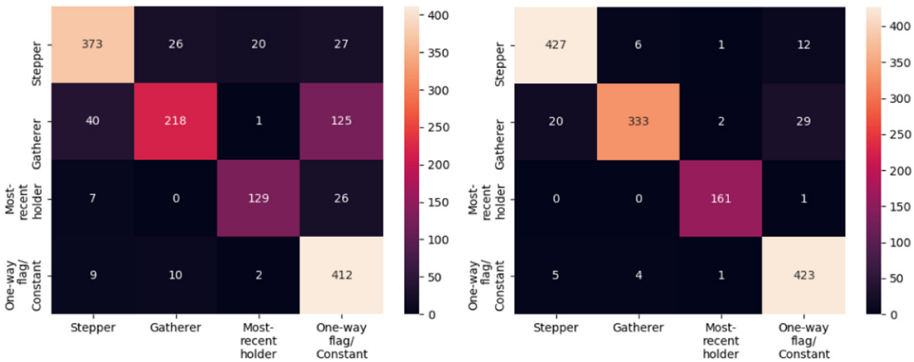


Fig. 6. The confusion matrix of SetTrace-1 (left) and SetTrace-5 (right) respectively. The x-axis indicates the predicted classes and the y-axis represents the ground-truth classes.

4.4 Model Training

The training of deep learning models is often tricky due to various issues such as overfitting and learning rate. Hyperparameter tuning had been carried out before the experiments to achieve model convergence and reasonable generation power. Figure 7 illustrates the insights of the training process for *SetTrace-5* through visualization of the loss function (i.e. error rate) and the accuracy against number of training iterations (i.e. epochs). At the end of training, both the loss function and the accuracy approached a steady level of low error rate.

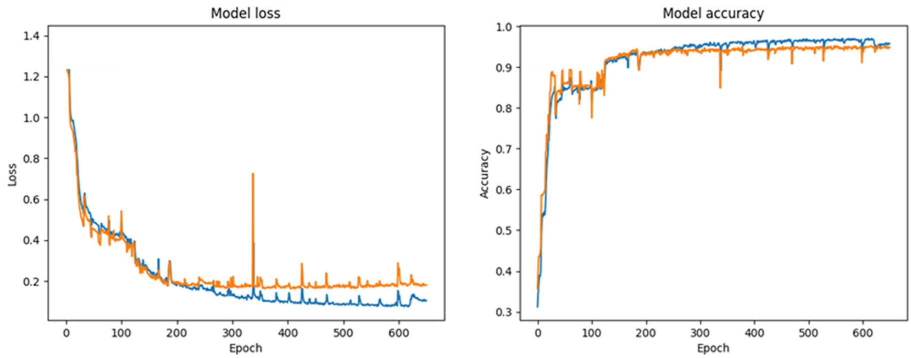


Fig. 7. The loss function and the model accuracy against the number of epochs in the training process. The blue curve represents the training set and the orange curve is the validation set.

Another training issue is the length of input variable traces. The characteristics of a variable may only be revealed in a longer trace for non-trivial programs. Long traces may overuse computer memory during training. A maximum length of 300 was found to balance the need for effective feature learning and the memory size limitation.

5 Discussion and Conclusion

A method for the recognition of variable roles in executable programs was developed and evaluated in an experimental study. The background, method and findings of the study have been described in the paper. The significance of the findings is to be described below.

The accurate prediction results showed that the role of a variable can be learned from execution-time variable traces. The set variable traces are superior to the single variable traces in manifesting the roles that involve interactions between two or more variables. RNN was found capable in learning the temporal features in the variable traces.

An interesting finding was that building two sub-models separately for the annotated variable and the context variables gave better performance. One probable explanation is that temporal sub-model for the annotated variable is basically the *SingleTrace* model, which was found capable of learning several variable roles. The traces of the context variables contain knowledge relevant to other roles. A suggestion for future work is to build several sub-models for capturing various facets of variable interactions and then to construct a classifier based on a concatenation of the sub-models.

A limitation of this study is the exclusion of some standard variable roles. In the 6 remaining variable roles proposed by Sajaniemi (2002), 3 of them, namely the *Most-wanted holder*, the *Follower*, and the *Temporary* roles, should be capably handled by the current method. The other three, the *Organizer*, the *Container*, and the *Walker* roles, must be acted by or with a data structure. A data structure can contain multiple value and the input shape does not fit the current architecture. A specialized input regulation may be a promising approach to resolve the problem.

The proposed method has characterized variable roles as execution-time patterns, and this is operationalized as variable traces. Obtaining variable traces can be tricky as it depends on the provision of the programming languages and the limitation of logging the update of variables. The technique developed for this study, for example, cannot capture the value before the first proper assignment statement for a variable. Variables of the *One-way Flag* role and the *Constant* role were found to produce the same type of traces. The two roles were therefore combined into a class in the study.

Automated recognition of variable roles can be an integral component of an intelligent tutor for computer programming. It can support asynchronous feedback for fixing errors in an online programming exercise as well as interactive advisement of variable use during the development of a programming solution. It must be stressed that the role of variables notion is an effective but a time-consuming pedagogy. Reviewing the variables manually is not trivial. In the classrooms, a teacher rarely has the time to review too many programs on the fly. An automated recognizer will make the pedagogy more practical in both classroom and online settings.

References

Ben-Ari, M., Sajaniemi, J.: Roles of variables as seen by CS educators. ACM SIGCSE Bull. **36**(3), 52–56 (2004)

Bishop, C., Johnson, C.G.: Roles of variables and program analysis. In: Proceedings of the 5th Finnish/Baltic Conference on Computer Science Education (2005)

Devlin, J., Uesato, J., Singh, R., Kohli, P.: Semantic Code Repair using Neuro-Symbolic Transformation Networks. arXiv, arXiv-1710 (2017)

Demyanova, Y., Veith, H., Zuleger, F.: On the concept of variable roles and its use in software analysis. In: Proceedings of the 2013 IEEE Formal Methods in Computer-Aided Design, pp. 226–230 (2013)

Gerdt, P., Sajaniemi, J.: A web-based service for the automatic detection of roles of variables. In: Proceedings of the 11th Annual SIGCSE Conference on Innovation and Technology in Computer Science Education, pp. 178–182 (2006)

Gupta, R., Pal, S., Kanade, A., Shevade, S.: DeepFix: fixing common C language errors by deep learning. In: Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence, pp. 1345–1351 (2017)

Hosanee, M., Rana, M.E.: A refined approach for understanding role of variables in elementary programming. J. Adv. Res. Dyn. Control Syst. **10**(11), 238–248 (2018)

Kant, N.: Recent advances in neural program synthesis (2018). arXiv preprint arXiv:1802.02353

Mou, L., Li, G., Zhang, L., Wang, T., Jin, Z.: Convolutional neural networks over tree structures for programming language processing. In: Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence, pp. 1287–1293 (2016)

Sajaniemi, J.: An empirical analysis of roles of variables in novice-level procedural programs. In: Proceedings IEEE 2002 Symposium on Human Centric Computing Languages and Environments, pp. 37–39. IEEE (2002)

Sajaniemi, J., Kuittinen, M.: An experiment on using roles of variables in teaching introductory programming. Comput. Sci. Educ. **15**(1), 59–82 (2005)

Sajaniemi, J., Navarro Prieto, R.: Roles of variables in experts' programming knowledge. In: Proceedings of the 17th Annual Workshop of the Psychology of Programming Interest Group (PPIG 2005) (2005)

Shi, N., Cui, W., Zhang, P., Sun, X.: Evaluating the effectiveness roles of variables in the novice programmers learning. J. Educ. Comput. Res. **56**(2), 181–201 (2018)

Wang, K., Singh, R., Su, Z.: Dynamic neural program embedding for program repair. arXiv preprint arXiv:1711.07163. (2017)

Author Index

Au, Oliver Tat Sheung 239

Bao, Haogang 227

Cai, Hongmei 112
Cerna, Miloslava 149
Chan, Chung-Shing 46
Chan, Ho-Yin 362
Chan, Pak-Kin 88
Chen, Gaowei 213
Chen, Kailiang 227
Chen, Wenyi 189
Cheng, Gary 281
Cheng, Victor Chi-Wa 88
Cheng, Xin 112
Cheung, Simon K. S. 271
Chui, Kwok Tai 136

Duan, Xialei 46

Fang, Sha 308
Fung, Yin-Chun 136

Gao, Ruixiang 176

Hao, Tianyong 323
Ho, Ching-hei 35
Hu, Jiahua 295
Hua, Xiaochen 23
Huang, Liyan 23, 258
Huang, Xiaoxiao 176
Huo, Liming 3

Ip, So-sum 35

Ji, Wenbin 201
Jia, Jiyou 100
Jittam, Piyachat 13

Kui, Chi-wang Daniel 35
Kwok, Chok-pang 35
Kwok, Jason Chun-Wai 136

Lam, Rechell Yee-Shun 349
Lam, Ringo Wai-Kit 88
Lau, Vincent King-Tang 88
Le, Huixiao 100
Lee, Lap-Kei 136, 239, 323
Lee, Victor Chung Sing 126
Lei, Shunwei 323
Leung, Tommy Cheuk-Hin 362
Li, Jinlin 323
Li, Kam Cheong 248
Li, Pan 201
Li, Wei 112
Li, Yanyan 227
Lin, Tzu-Han 164
Liu, Xiaobin 75
Liu, Yirao 176
Lo, Chung Kwan 213
Lui, Andrew Kwok-Fai 362
Lui, Tsz-Tik 362

Mahardthai, Siwa 13
Mai, Ziying 3
Mei, Hu 176

Ng, Kwan-Keung 349
Ng, Ricky Yuk-Kwan 349
Ng, Sin-chun 35
Ng, Wing Shui 281

Peng, Yu 227
Poon, Chung Keung 126
Pu, Qiuyu 61

Rao, Yanghui 295

Shen, Yingshan 323
Shi, Yinghui 61
Song, Yu 323
Sriwattanarothai, Namkang 13
Su, You 227
Sun, Ning-wei 271
Sun-Lin, Hong-Zheng 164

Tam, Chi Man 308
Tang, Chung Man 126

U, Leong Hou 136

Wang, Fu Lee 239, 295
Wang, Jiaying 23
Wang, Jingjing 239
Wang, Meihua 201
Wang, Yilu 258
Wong, Billy Tak-Ming 248
Wong, IpKin Anthony 46
Wong, Tak-Lam 126
Wu, Qingyuan 295
Wu, Xiaohong 308

Xiao, Jun 164
Xie, Haoran 239, 281, 295
Xie, Yanna 258

Yan, Ning 337
Yang, Harrison Hao 61
Yang, Huiyun 61
Yang, Xuesong 112
Yao, Xiao 3
Yodyingyong, Supan 13
Yu, Yuen Tak 126
Yuan, Yuting 75

Zeng, Jielin 176
Zhan, Tian-Jie 88
Zhan, Zehui 3, 176
Zhang, Jinbin 201
Zheng, Yunxiang 3
Zou, Di 239