Educational Technology & Society has Impact Factor 4.020 and
5-Year impact factor 4.865 according to Thomson Scientific 2022 Journal Citations Report.

http://www.j-ets.net/

Jul. 2023

Educational Technology & Society

vol.26
no.3

Educational Technology & Society

Published by International Forum of Educational Technology & Society
Hosted by National Taiwan Normal University, Taiwan

Volume 26 Issue 3

July 2023

ISSN: 1436-4522 (online)
ISSN: 1176-3647 (print)
DOI: 10.30191/ETS
http://www.j-ets.net/

Educational Technology & Society

An International Journal

Aims and Scope

Educational Technology & Society (ET&S) is an open-access academic journal published quarterly (January, April, July, and October) since October 1998. *ET&S* has achieved its purposes of providing an international forum for open access scientific dialogue for developers, educators and researchers to foster the development of research in educational technology. Thanks to all the Authors, Reviewers and Readers, the journal has enjoyed tremendous success.

ET&S has established a solid and stable editorial office where the Editors-in-Chief aims to promote innovative educational technology research based on empirical inquiries to echo the pedagogical essentials of learning in the real world—lifelong learning, competency-orientation, and multimodal literacy in the 21st century.

ET&S publishes research that well bridges the pedagogy and practice in advanced technology for evidence-based and meaningful educational application. The focus of *ET&S* is not only technology per se, but rather issues related to the process continuum of learning, teaching, and assessment and how they are affected or enhanced using technologies. The empirical research about how technology can be used to overcome the existing problems in the frontline of local education with findings that can be applied to the global spectrum is also welcome. However, papers with only descriptions of the results obtained from self-report surveys without systematic or empirical data or any analysis on learning outcomes or processes are not favorable to be included in *ET&S*.

ET&S publishes research that fulfills any of the following description:

- **Evidence-based Research:** Continuous research that is conducted within a sufficient amount of time such that the effectiveness of the research intervention in education can be evaluated and demonstrated convincingly via empirical methods. Preferably, research experience, outcome, and results for a semester long of adopting educational technologies; or an evaluation/experiment that is conducted multiple times and spans over several semesters.
- **Emerging Technology:** New and emerging technology used in education
- **Special Domain Research:** Research with specific participants using technology in education, for example, those with privacy issues in special education or those of younger ages like infants and toddlers, who may not form a large pool of participants

ISSN 1436-4522 (online) and 1176-3647 (print). DOI 10.30191/ETS.

Publisher
International Forum of Educational Technology & Society

Host
National Taiwan Normal University, Taiwan

Editorial Office
c/o Chair Professor Nian-Shing Chen, Institute for Research Excellence in Learning Sciences, Program of Learning Sciences, National Taiwan Normal University, No.162, Sec. 1, Heping E. Rd., Da-an Dist., Taipei City 10610, Taiwan.

Supporting Organizations
University of North Texas, USA
University of Piraeus, Greece

Website Maintenance
Institute for Research Excellence in Learning Sciences, National Taiwan Normal University, Taiwan

Abstracting and Indexing
Educational Technology & Society is abstracted/indexed in Social Science Citation Index, Scopus, ACM Guide to Computing Literature, airiti, Australian DEST Register of Refereed Journals, Computing Reviews, Current Contents/Social & Behavioral Sciences, DBLP, DOAJ, Educational Administration Abstracts, Educational Research Abstracts, Educational Technology Abstracts, Elsevier Bibliographic Databases, ERIC Clearinghouse on Information & Technology, Inspec, ISI Alerting Services, JSTOR, PsycINFO, Social Scisearch, Technical Education & Training Abstracts, and VOCED.

Guidelines for authors
Submissions are invited in the following categories:

- Peer reviewed publications: Full length articles (up to 8,000 words excluding References and Appendices)
- Special Issue publications

All peer review publications will be refereed in double-blind review process by at least two international reviewers with expertise in the relevant subject area.

For detailed information on how to format your submissions, please see: https://www.j-ets.net/author_guide

For Special Issue Proposal submission, please see: https://www.j-ets.net/journal_info/special-issue-proposals

Submission procedure
All submissions must be uploaded through our online management system (http://www.j-ets.net). Do note that all manuscripts must comply with requirements stated in the Authors Guidelines.

Authors, submitting articles for a particular special issue, should send their submissions according to the channel specified in the Call for Paper of the special issue.

All submissions should be in electronic form. Authors will receive an email acknowledgement of their submission.

The preferred formats for submission are Word document, and not in any other word-processing or desktop-publishing formats. Please place figures and tables in their respective format in the anonymous manuscript along with all appendices (if any).

Please provide following details with each submission in a separate file (i.e., Title Page):

- Author(s) full name(s) including title(s),
- Name of corresponding author,
- Job title(s),
- Organisation(s),
- Full contact details of ALL authors including email address, postal address, telephone and fax numbers.

In case of difficulties, please contact journal.ets@gmail.com.

Educational Technology & Society

Volume 26 Number 3 2023

Table of Contents

Editorial Note

Contextualized Multimodal Language Learning 1–4
Meei-Ling Liaw and Hsin-I Chen

Special Issue Articles

Using an AI-Based Object Detection Translation Application for English Vocabulary Learning 5–20
Pei-Lin Liu and Chiu-Jung Chen

Investigating Adolescents' Participation Trajectories in a Collaborative Multimodal Composing Learning Environment 21–36
Shiyan Jiang

Effects of Automatic Speech Recognition Technology on EFL Learners' Willingness to Communicate and Interactional Features 37–52
Michael Yi-Chao Jiang, Morris Siu-Yung Jong, Wilfred Wing-Fat Lau, Ching-Sing Chai and Na Wu

Cluster Analysis of Hong Kong Students' Self-Regulated Learning in Contextualized Multimodal Language Learning 53–68
Lucas Kohnke, Dennis Foung and Julia Chen

Exploring Students' Experiences of Using Multimodal CMC Tasks for English Communication: A Case with Instagram 69–83
Muntaha Muntaha, Julian Chen and Toni Dobinson

Improving Language Learning Activity Design through Identifying Learning Difficulties in a Platform Using Educational Robots and IoT-based Tangible Objects 84–100
Yueh-hui Vanessa Chiang, Ya-Wen Cheng and Nian-Shing Chen

Exploring Multiliteracy of Pre-Service Language Teachers through Spherical Video-Based Virtual Reality 101–114
Rae Ping Lin and Wen-Chi Vivian Wu

Full Length Articles

Advancing University EFL Students' Argumentative Essay Writing Performance through Knowledge-Building-based Holistic Instruction 115–128
Ying-Tien Wu and Li-Jen Wang

The Demographics of Student Device Ownership: An Examination of the Personal Computing Ecosystems of Students in Higher Education 129–140
Rob Elliott

Students' Learning Styles and Preferences in a Gamification-enhanced Partially Flipped Classroom: A Q-Methodology Study 141–154
Liwen Chen

Implementing Unplugged CS and Use-Modify-Create to Develop Student Computational Thinking Skills: – A Nationwide Implementation in Colombia 155–175
Camilo Vieira, Ricardo L. Gómez, Margarita Gómez, Michael Canu and Mauricio Duque

Exploring Potential Factors to Students' Computational Thinking: Interactions between Gender and ICT-resource Differences in Taiwanese Junior High Schools 176–189
Silvia Wen-Yu Lee, Jyh-Chong Liang, Chung-Yuan Hsu, Francis Pingfan Chien and Meng-Jung Tsai

Learning Performances towards the BookRoll E-Book System for Flipped Classrooms in Software Engineering Education 190–202
Yen-Ting Lin

Liaw, M.-L., & Chen, H.-I. (2023). Guest Editorial: Contextualized Multimodal Language Learning. *Educational Technology & Society*, *26*(3), 1-4. https://doi.org/10.30191/ETS.202307_26(3).0001

Guest Editorial: Contextualized Multimodal Language Learning

Meei-Ling Liaw[1*] and Hsin-I Chen[2]

[1]National Taichung University of Education, Taichung, Taiwan // [2]National Taipei University of Technology, Taipei, Taiwan // meeilingliaw@gmail.com // hichen@mail.ntut.edu.tw

[*]Corresponding author

ABSTRACT: We live in an era of digitally accessible multimodality for various purposes and practices. Researchers and educators agree that multimodal literacies are essential by human beings to communicate, work, and thrive in the global world of the 21st century (Gee, 2003; Jewitt & Kress, 2003; New London Group, 1996). Along with this need, teachers need to be aware of "multimodal possibilities" (Lotherington & Jenson, 2011, p. 227) and their ramifications for teaching and learning. In second/foreign language education, multimodality has become even more central than ever. The interconnectedness among learning contexts, digital tools and materials, and learners is dynamic, multi-faceted, and, more importantly, awaits further exploration so language teachers and learners can transform the understandings into effective pedagogical practices. In this special issue, we present seven research efforts contributing to moving toward this goal. Under the overarching theme of contextualized multimodal language learning, the studies tackle issues in theoretical perspectives, methodological choices, educational contexts, and applications of innovative technological tools. Collectively, the studies revealed positive pedagogical values for language teachers of different educational contexts to enhance the learning experiences of different age groups by creatively taking advantage of multiple modes of knowing and meaning-making.

Keywords: Multimodal, Contextualized language learning, Digital tools, Language teaching

1. Introduction

All interactions are multimodal by nature (Norris, 2004). Humans use various modes to represent meanings and exchange ideas in communication. Multimodal language learning is coined upon the fact that humans integrate different modes, including audio, textual, gestural, visual, and spatial resources, to learn languages. In second and foreign language learning, skills in interpreting mediated modalities are therefore necessary for language learners, especially when communication takes place in cross-boundary and cross-cultural contexts. Digital technologies allow the combined use of texts, images, audio, videos, and multimedia, at the same time, further intensify these so-called "multimodal possibilities" (Lotherington & Jenson, 2011, p. 227). Although multimodality provides great potential for language learning when digital tools are involved, the multitude of possible multimodal compositions and their impacts on learning and interaction can easily overwhelm learners of an additional language (Abrams, 2016; Hampel & Hauck, 2006). With the affordances of digital tools, how multimodal affordances affect language learners' meaning-making and language learning cognitively, affectively, and socio-culturally in technology-enhanced language learning (TELL) contexts thus deserve in-depth explorations.

As Bax (2003) foreseen the technology to progress towards normalization almost 20 years ago, technology nowadays has become invisible, embedded in everyday practice, and truly integrated into our lives. Language learners are engaged in technology-enhanced practices in formal schooling settings and digital wilds that go beyond contexts within educational systems (Sauro & Zourou, 2019). Language learning in digital contexts is rather open, dynamic, multi-faceted, and unpredictable. To have a comprehensive understanding of learners' multimodal practice and performance in TELL activities, it is essential to contextualize learners' multimodal experiences and examine the interrelationship between the multimodal potentialities of diverse learning settings, the mediated human activity within those settings, and the characteristics of the learners. The interconnectedness among contexts, TELL activities, and learners, are complex and multi-faceted. Thus, some questions arise: How does the multimodality feature affect language learning and use in different learning contexts? How to leverage technologies to foster learners' multimodal language learning? How does introducing and integrating emerging technologies enable learners and teachers to be exposed to and/or create contextualized multimodal language learning experiences? The seven articles included in this special issue tackle these questions from different perspectives concerning different aspects of (de)contextualization of learning practices.

2. In this issue

To address the aforementioned questions, this special issue openly invited contributions from researchers engaging in scholarly endeavors focusing on multimodal uses and their effects on second and foreign language learning. The seven studies in this issue represent a whole array of such practices, ranging from various theoretical perspectives, methodological choices, and pedagogical contexts to applications of innovative technological tools. We hope the publication of this special issue will help promote reciprocal dialogues facilitating the expansions of our knowledge of the theoretical underpinnings and pedagogical practicalities of contextualized multimodal language learning (CMLL) in various digital spaces. In the following, we briefly describe each article in this issue:

The first article of this special issue, entitled "Using an AI-based object detection translation application for English vocabulary learning," described a learning system featuring the AI-based ODT app developed by the researchers and its effects on elementary school students' English vocabulary learning. The study found that the students who learned using the AI-based ODT app technology outperformed those who learned using the non-AI technique in terms of test scores. The researchers thus suggest that the AI-based ODT app could be useful as a teaching aid for young children.

The second article of this special issue, "Toward broadening participation: Investigation adolescents' participation trajectories in a collaborative multimodal composing learning environment," is a multiple case study investigating fifth to eighth graders' participation trajectories when engaged in creating multimodal science fiction stories in small groups. The author's analyses of multiple data sources revealed that the participants could use multiple modes to move across forms for interdisciplinary meaning-making and demonstrate their expertise as knowledge producers. Based on the findings, the research calls for further studies on how adolescents changed the form and degree of participation in integrated STEM learning environments.

The third article of this special issue, entitled "Effects of automatic speech recognition technology on Chinese EFL learners' willingness to communicate," examined the effects of automatic speech recognition (ASR) technology on university student' willingness to communicate (WTC) in oral English. The quasi-experimental study involved 160 students using ASR technology with a flipped classroom approach. The findings showed that the target group had significantly higher post-intervention WTC scores with teacher and class than the control group students. However, the development trajectories of the students' interactional features indicated that the use of the ASR-based technology might have only exerted a limited effect on the participants' in-class peer interaction due to the short period of use of ASR and cultural factors. The researchers concluded that when applying technologies to designing activities for students' WTC, cultural factors may need to be taken into consideration.

The fourth article of this special issue, entitled "Cluster analysis of Hong Kong students' self-regulated learning (SRL) in contextualized multimodal language learning," investigated the relationship between self-regulated learning and academic success. The researchers employed cluster analyses on the SRL behaviors of university students in Hong Kong as they worked on the content in an in-house developed multimodal learning package. Statistical procedures were used to explore the differences between clusters. Clusters of students who differed distinctly in their SRL behaviors were identified. The study also revealed that good mastery of SRL was strongly related to course outcomes in the contextualized language course possessing certain assessment components. Based on the findings, the researchers emphasize the importance of developing personalized instructions to motivate, stimulate, and foster SRL.

The fifth article of this special issue, "Exploring students' experiences of using multimodal CMC tasks: A case with Instagram," explored Indonesian university students' learning experiences with multimodal CMC tasks through Instagram. Analyses of pre- and post-study surveys, journal reflections, and interviews revealed positive student responses to the learning approach. Students considered the CMC tasks fun and enjoyable, and the paralinguistic features afforded by Instagram helped them communicate more effectively.

The sixth article of this special issue, "Improving language learning activity design through identifying learning difficulties in a platform using educational robots and IoT-based tangible objects," aimed at understanding the obstacles a group of elementary school students faced when taking part in learning activities supported by robots and IoT-based tangible objects. The researchers analyzed the video recordings of the participants' learning process and identified categories of obstacles preventing learners from completing the tasks and the causes of the obstacles. Based on the findings, the researchers offered instructional guidelines for designing learning activities using the robot and IoT-based tangible objects.

The seventh article of this special, "Exploring multiliteracy of pre-service language teachers through spherical video-based virtual reality," reported how a group of Taiwanese pre-service teachers' multiliteracies developed through utilizing spherical video-based virtual reality (SVVR) tools to design teaching materials in an English as a Foreign Language (EFL) context. The researchers collected video recordings of the participants' presentations about their final SVVR artifacts and semi-structured online interviews. The findings from video recording analyses revealed that, through developing SVVR teaching materials, the per-service EFL teachers learned to compose multimodal lessons, concretized the intangible context for learning, and viewed the virtual space as a mode for teaching and learning. The interview results echoed such positive findings.

3. Conclusion

Context is a multi-faceted and multi-dimensional concept in language education. It covers different aspects, including educational settings, learning modes, sociocultural relations, and media modalities (Luckin, 2010). Language learners' learning experiences can be perceived as transitions between settings defined by these aspects (Glahn & Gruber, 2020). Contextualized language learning implies that all language learning is situated in the context of real-world activities in which learners are encouraged to take active roles in meaning-making and problem-solving. Contextualized language learning places learners at the center of learning and engages learners in authentic contexts that are relevant to learners. Contextualized multimodal language learning further emphasizes the multimodality and multiplicity of communicative modes available in situated learning practices. With the accessibility and affordances of emerging technologies in the digital age, multimodality and CMLL have become even more central to communication and language learning and teaching (Dressman, 2020).

The studies in this special issue revealed positive pedagogical values for language teachers to enhance students' learning experiences by creatively taking advantage of multiple modes of knowing and meaning-making. Such positive pedagogical effects occurred in elementary school classrooms as well as higher education institutions. CMLL can be deconceptualized from the explorations of the applications of different emerging technologies, including social media, virtual reality, Automatic Speech Recognition, robots, AI-based object detection translation; different constructs of learning including willingness to communicate, participatory, learner attitude, teacher attitude; different target language skills such as vocabulary acquisition, writing, and speaking; different research methodologies and inquiries including quasi-experimental study, qualitative case study, statistical analyses, exploratory, and mixed method; different learner population that includes young learners, adolescent learners, adult learners, and pre-service teachers. It is hoped that through the collection of seven studies, this special issue underpins the multiplicity, dynamics, and complexity of theoretical groundings and pedagogical implications of contextualized multimodal language learning for second/foreign language pedagogy.

References

Abrams, Z. I. (2016). Possibilities and challenges of learning German in a multimodal environment: A Case study. *ReCALL, 28*(3), 343-363.

Bax, S. (2003). CALL—Past, present and future. *System, 31*, 13-28.

Dressman, M. (2020). Multimodality and language learning. In M. Dressman & R. W. Sadler (Eds.), *The Handbook of Informal Language Learning* (pp. 39-55). Wiley-Blackwell. https://doi.org/10.1002/9781119472384

Gee, J. P. (2003). *What video games have to teach us about learning and literacy*. Palgrave Macmillian

Glahn, C., & Gruber, M. R. (2020). Designing for context-aware and contextualized learning. In S. Yu, M. Ally, & A. Tsinakos (Eds.), *Emerging Technologies and Pedagogies in the Curriculum. Bridging Human and Machine: Future Education with Intelligence* (pp. 21–40). Springer. https://doi.org/10.1007/978-981-15-0618-5_2

Hampel, R., & Hauck, M. (2006). Computer-mediated language learning: Making meaning in multimodal virtual learning spaces. *The JALT CALL Journal, 2*(2), 3-18.

Jewitt, C., & Kress, G. (Eds.). (2003). *Multimodal literacy.* Peter Lang.

Lotherington, H., & Jenson, J. (2011). Teaching multimodal and digital literacy in L2 settings: New literacies, new basics, new pedagogies. *Annual Review of Applied Linguistics, 31*, 226-246.

Luckin, R. (2010). *Re-designing learning contexts. Technology-rich, learner centered ecologies*. Routledge.

New London Group. (1996). A Pedagogy of multiliteracies: Designing social futures. *Harvard Educational Review, 66*, 60-92.

Norris, S. (2004). *Analyzing multimodal interaction: A Methodological framework*. Routledge.

Sauro, S., & Zourou, K. (2019). What are the digital wilds? *Language Learning & Technology*, *23*(1), 1-7.

Liu, P.-L., & Chen, C.-J. (2023). Using an AI-Based Object Detection Translation Application for English Vocabulary Learning. *Educational Technology & Society*, *26*(3), 5-20. https://doi.org/10.30191/ETS.202307_26(3).0002

Using an AI-Based Object Detection Translation Application for English Vocabulary Learning

Pei-Lin Liu[1] and Chiu-Jung Chen[2*]
[1]National Chia-Yi University, Taiwan // [2]National Chia-Yi University, Taiwan // peilin@mail.ncyu.edu.tw // chenc@mail.ncyu.edu.tw
[*]Corresponding author

ABSTRACT: This study aimed to examine the effects of an AI-based object detection translation (AI-based ODT) application (app) on EFL students' vocabulary learning. We developed a system that utilized strategies to facilitate learners' vocabulary learning. The app applied dual code theory to present the objects in picture, word, and pronunciation formats. Seventy-two elementary school students were divided into lower-ability and higher-ability groups according to their English proficiency, and were then randomly assigned to the control and experimental conditions based on their ability. The learners' learning performance in the control and experimental conditions was compared using a pre-test–post-test design. Through two-way ANOVA analysis, we observed that in the experimental group the higher-ability students benefited more from the AI-based ODT app technology than did the lower-ability students. This significant difference could be taken as evidence of the positive effect of the AI-based ODT app technology, particularly for higher-ability students.

Keywords: Artificial intelligence, Dual code theory, English as a foreign language (EFL), Object detection translation, Vocabulary learning

1. Introduction

1.1. Background and problems of the study

Learning foreign languages can be a challenging but rewarding process for young children. Learning a language in countries where that language is not generally spoken can be difficult. Among the language learning skills, the use of a foreign language primarily relies on the use of vocabulary as the building blocks of a language. Without vocabulary, people cannot communicate or continue to develop their language acquisition (Ramos & Dario, 2015).

As a result, improving vocabulary learning is essential and a priority in the language learning process in order to learn a language well (Tanaka, 2017). Creating an impactful educational setting is crucial for increasing young children's motivation for learning (Weiland & Yoshikawa, 2013). Traditionally, English is taught to children using songs, textbook exercises, nursery rhymes, and storybook reading, and a majority of learners apply the rote learning strategy for vocabulary memorization based on repetition (Nation, 2013). However, there are some common difficulties associated with the rote learning of vocabulary, including the learners easily forgetting words and having difficulty recalling them (Chuo & Yen, 2014; Wu & Huang, 2017).

In order to overcome the limitations of rote learning, the provision of effective learning strategies and tools to help learners improve their memorization of vocabulary is important. Today's children are raised in a technologically advanced society, which has influenced how they learn in comparison to previous generations. Lessons which incorporate technology can help to shape the teaching and learning process to be more inventive. It is critical to investigate which technologies can be used in the classroom to engage young non-native-speaking pupils in the learning materials and to encourage them to practice English in order to enhance their foreign language learning skills (Dalim et al., 2020).

1.2. Purpose and scope of this study

This study explored the potential of combining an artificial intelligence (AI) based object detection translation (ODT) application (app) and dual code theory design which presented the objects in picture, text, and pronunciation formats. The AI-based ODT app was adopted as an innovative way to teach basic English vocabulary to young non-native-speaking learners. AI is the general term for the science of artificial intelligence. It uses computers to simulate human intelligent behaviors and it trains computers to learn human behaviors such as learning, judgment, and decision-making (Yigitcanlar et al., 2020; Zhang & Lu, 2021). Object detection is a subset of AI; it uses deep learning to provide a fast and accurate means to predict the location of an object in an image. Deep learning is a powerful machine learning technique that uses multi-layered AI networks in which the object detector automatically learns image features required for detection accuracy in tasks. For example, Google Translate (see Figure 1) allows learners to explore the world in the language that they are familiar with by just pointing the camera lens at the foreign text (Gu, 2019; Liu, 2018). The function of Google Translate is to use AI to train computers through machine learning and deep learning technologies in order to predict the most likely words, thus providing stronger and more accurate translations (Liu, 2018).

Figure 1. The Google Translate app (Fu, 2016)

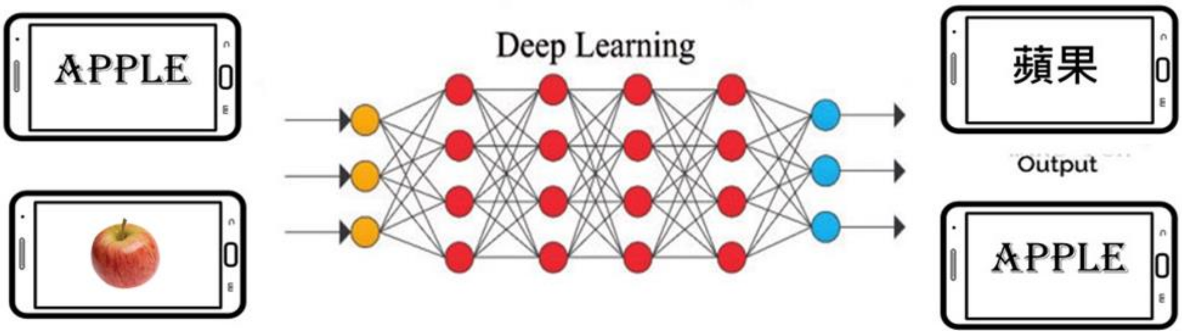
The Google Translate app has been updated to include 60 new languages. When a camera image is sent, the app must first locate the letters in the image. It must filter out the background items and focus on the words that it can identify. Second, the app must be able to identify each letter. Deep learning comes into play here. The third step is to look up the known letters in a dictionary to check what their particular combination means (see Figure 2) (Good, 2015; Vincent, 2019).

Figure 2. Step-by-step process of how the Google Translate app works (Good, 2015)

Although the Google Translate app launched the Instant Camera Translation function for mobile phones in 2014, thus far, there have been few empirical studies on its application in English learning. Most of the related studies have focused on the technical introduction and calculations to improve image recognition (e.g., Chatterjee & Bhattacharjee, 2020; de Carvalho et al., 2018; Giovany et al., 2017; Kong et al., 2015; Mezgec et al., 2019; Xie et al., 2020) and text recognition (Lee et al., 2017; Yousef et al., 2020; Zhong et al., 2018). Both Giovany's et al. (2017) and Kong's et al. (2015) studies concentrated on how to improve the recognition of food images, with a focus on how to remove unnecessary information due to the cluttered background on the menu, and how to improve recognition of the menu text.

In terms of the current computer technology, recognizing "text" is simple for a computer. Because of the fixed writing method of text or numbers, the computer can analyze the characters using a database. However, it is relatively difficult to detect "objects." An object involves different compositions and details, the identification of which requires more complicated calculations by the computer. This is also one of the reasons why Google Translate can only support text recognition and translation functions for objects, but is currently unable to support an object detection translation function (see Figure 3).

Figure 3. Differences between the text-to-text translation app and the object detection translation app



In the present study, we developed a learning system featuring the AI-based ODT app (see Figure 4) based on the dual-coding theory (DCT). The dual-coding theory assumes that information is encoded using both visual and verbal forms. In the human mind, both visual and verbal information are processed separately via different channels. By integrating both visual and verbal channels simultaneously, it will be easier for learners to recall the particular memory in the future (Clark & Paivio, 1991; Kanellopoulou et al., 2019; Kassim, 2018). In the study of English vocabulary, it has been proven that the dual-coding theory can provide learners with the ability to organize and classify words through associations and present the relevance of words with specific pictures, not only allowing learners to memorize these words quickly but also helping them to remember the words for a long time (Liu, 2016; Liu et al., 2020).

Figure 4. Schematic representation of the ODT app developed by the researchers

1.3. Benefits of using the ODT app with AI, and the research question

The combination of the AI-based ODT app and the DCT design in this study enabled the presentation of objects in picture, text, and pronunciation formats. Learners took photographs of objects, and the AI-based ODT app recognized and translated the authentic objects into English/Chinese words with their relevant pronunciation. There are two main benefits of using ODT with the application of AI. First, it is expected that well-validated AI tools can be used to improve the accuracy and reproducibility of vocabulary learning. Moreover, AI systems reproducibly yield the same results when provided with similar inputs, do not suffer from fatigue, and excel at finding patterns in large amounts of data. Second, AI tools can augment the efficiency of object detection for language learning (Steiner et al., 2021).

Driven by the dual needs for improved accuracy and efficiency, as well as by rapid improvements in technology, there are now a growing number of research articles describing promising applications of AI across a wide variety of tasks. However, despite these fundamental advances, there are exceptionally few practical demonstrations of the integration of AI into language learning.

The aim of this study was therefore to explore the potential of an AI-based ODT app for teaching the English names of fruit and vegetables to children who do not speak English as their first language. This study aimed to answer the following research question: Is there a significant difference in the learning gains of children with different abilities due to the teaching platforms in which they experienced second language vocabulary learning?

1.4. Contributions of this study

The main contributions of the study include insights into how object recognition translation affects children's interaction with an AI-based application, and influence their learning experience. The more that is known about the effect of taking pictures and AI on young non-native children's English language learning, the more effective learning approaches that can be developed.

2. Literature review

2.1. Technology-enhanced language learning

Learning and teaching processes have entirely changed as a result of the availability of new interactive computer technologies (Burston, 2014; Eslami & Ahmadi, 2019), leading to the development of several fields of study such as Computer-Assisted Language Learning, Mobile-Assisted Language Learning, and Computer-Mediated Communication. These distinct but overlapping fields of study share a focus on using technology as an assistive tool or mediator to enhance the teaching and learning of a second or foreign language (L2) (Chun, 2016). In most cases, technology has been used to assist students in accomplishing a variety of language learning objectives such as listening comprehension (Ramírez Verdugo & Alonso Belmonte, 2007), reading comprehension (Dreyer & Nel, 2003), and vocabulary acquisition (Oberg, 2011). In terms of terminology, Computer-Assisted Language Learning (CALL) is a well-established field with a long history, and is often used as a catch-all phrase for research on computer use in L2 settings (Bax, 2003; Bax, 2011). In recent years, the term Technology-Enhanced Language Learning (TELL) has often been used interchangeably with or even in preference to CALL (Hubbard, 2013; Walker & White, 2013), as TELL suggests a more inclusive sense of technology interventions enabled by a wide variety of technology devices and software applications. As a result, TELL is used in this article to refer to the employment of technology in various forms in the application domain of L2 education (Chang & Hung, 2019).

Many scholars have discussed the advantages of mobile devices in education, including broadening learning outside the conventional classroom (Wu, 2016), allowing convenient access to learning materials (Kaliisa et al., 2019), flexibility in the time and location of study (Loewen et al., 2019), and creating more interaction and communication between teachers and students (Wu, 2014). The related studies include the use of mobile devices for vocabulary learning (Asmana & Arifani, 2021; Katemba, 2021), reading comprehension (Alnujaidi, 2021; Rahimi & Babaei, 2021), sentence construction (Purgina et al., 2017), listening (Islam & Hasan, 2020; Kamasak et al., 2021), speaking skills (Almadhady et al., 2019; García Botero et al., 2019; Lutfi, 2020; Xu, 2020), writing (Gharehblagh & Nasri, 2020; Krisbiantoro, & Pujiani, 2021), and grammatical development (Ghorbani & Ebadi, 2020). With the development of smartphones, an increasing number of applications have been developed to offer a variety of functions that can assist with English learning. The use of technologies for language learning has become nearly ubiquitous.

New technologies (e.g., AI, virtual reality, augmented reality, or wearable technologies) are increasingly available (Shadiev et al., 2019). Furthermore, emerging technologies are maturing and are very promising for use in language learning and instruction. Studies have shown that technology can promote the learning performance of language learners, increase learning motivation, and provide them with more efficient means of language learning (Jin, 2018; Shadiev & Huang, 2020).

Shadiev and Yang (2020) reviewed 398 research articles on technology-enhanced language learning and teaching from 2014 to 2019. The types of technology used in the articles we reviewed are listed in Table 1 (adopted from Shadiev & Yang, 2020). In these articles, the most used technologies were games (n = 49) and online videos (n = 37). There are, however, few studies that have explored the use of new technologies (e.g., AI n = 0, VR n = 19, AR n = 3, wearable technology n = 2)

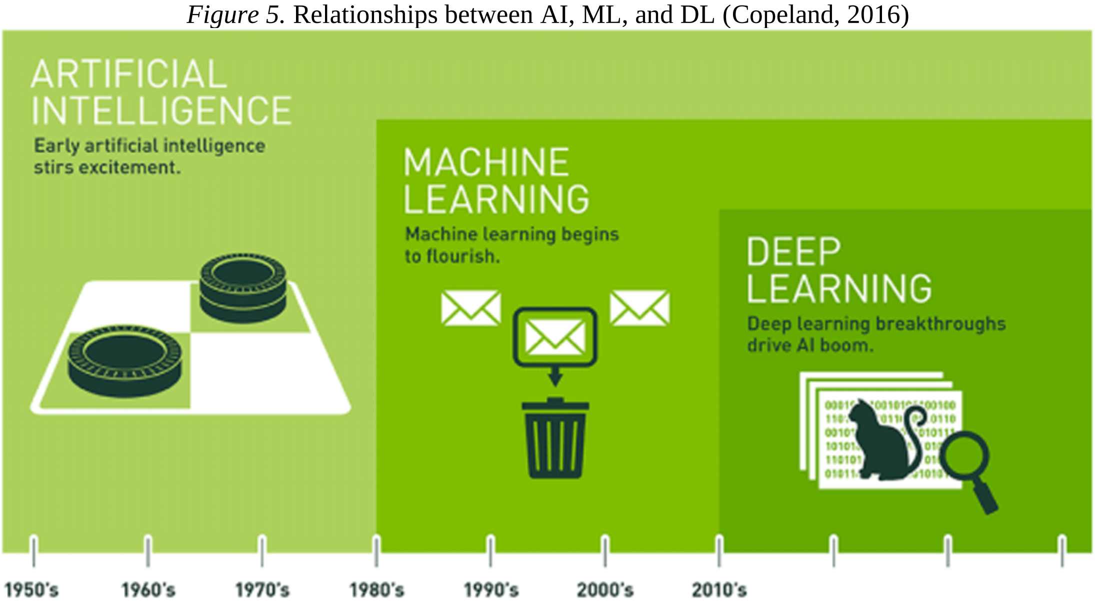
Table 1. Technologies used in language learning and teaching from 2014 to 2019

Technology	Year						Total
	2014	2015	2016	2017	2018	2019	
Game	6	7	6	11	14	5	49
Online video	5	7	4	6	10	5	37
Collaborative writing	3	9	8	5	3	7	35
Corpus	8	5	9	5	2	3	32
Instant messaging	2	4	4	11	5	5	31
Automated feedback	6	4	4	4	5	6	29
Social networking	1	6	4	10	7	1	29
Websites and digital resources	3	2	5	6	3	6	25
Virtual reality	0	5	2	4	6	2	19
Speech recognition	0	3	5	4	3	3	18
Electronic gloss or annotation	3	1	0	3	2	1	10
E-books	0	0	3	3	2	1	9
Electronic dictionary	2	0	3	1	1	1	8
Intelligent tutoring system	1	1	1	1	0	2	6
Voice recording	2	0	0	1	2	1	6
Augmented Reality	0	0	0	2	1	0	3
Robots	0	1	1	0	1	0	3
Clicker	0	0	0	0	1	2	3
Wearable devices	0	0	1	0	1	0	2
Course management system	2	0	0	0	0	0	2
Digital library	1	0	0	0	0	0	1
White board	1	0	0	0	0	0	1
Unidentified technology	7	6	9	5	9	12	48

2.2. AI and language acquisition

Artificial intelligence (AI) is a branch of computer science that focuses on the creation of intelligent machines that think and work like humans (Xie et al., 2021). Its goal is to replicate human intelligence in machines that are programmed to think and act like humans. AI plays an important role in the technology that supports daily social life. Machine learning (ML) is an AI system that can learn on its own by using an algorithm (Rahimy, 2018). Deep learning (DL) is a type of machine learning that is used to analyze large amounts of data by using a cascade of multilayered convolutional neural networks (CNN) (Lee et al., 2021; Litjens et al., 2017; Shen et al., 2017) (see Figure 5).

Figure 5. Relationships between AI, ML, and DL (Copeland, 2016)



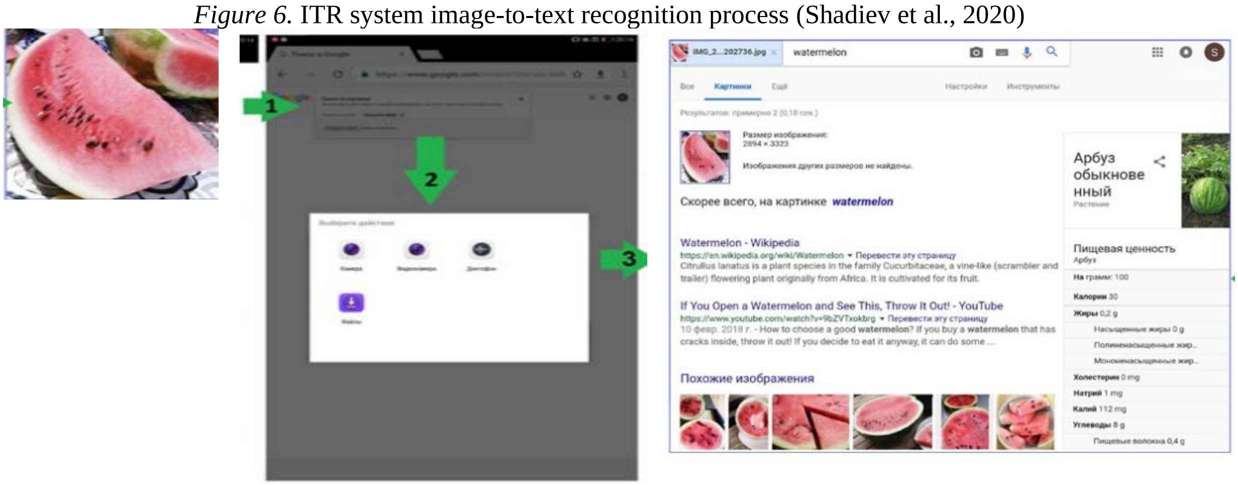
The emerging AI technologies in teaching and studying are used to provide more customized, versatile, inclusive, and engaging learning, as well as to automate everyday learning activities via automated evaluation and feedback (Gulson et al., 2018; Luckin et al., 2016). Theoretically, AI can assist parents in optimizing their children’s early linguistic development, as well as teachers in choosing resources, planning courses, increasing attendance, and delivering customized instruction to their students (Porayska-Pomsta, 2016). In its early stages, AI in education generally referred to intelligent tutoring systems that aimed to solve problems such as automatically improving the operator performance (e.g., Chai et al., 2021; Hwang, 2003; Ross, 1987). Nowadays, AI refers to the use of large amounts of data to complete complex tasks.

DL has sparked a surge of interest in business and science over the last decade, revolutionizing the field of machine learning by achieving state-of-the-art results in perception tasks such as image and speech recognition (Goodfellow et al., 2016). Recently, DL has obtained the top results in a wide range of computer vision issues, and has proved to be particularly effective for image recognition (Mezgec et al., 2019), a visual deep learning model, in which a convolutional neural network (CNN) is applied. A CNN is made up of thousands of individual neurons that can perform complex tasks, such as pixel intensity-based image recognition and classification (Rahimy, 2018). In other words, deep learning techniques allow for this automatic learning by absorbing large quantities of unstructured data such as text, images, or video (Salloum et al., 2020).

To the best of our knowledge, thus far, there have been only a few studies on the application of AI to language learning (i.e., Hasnine et al., 2019; Hasnine et al., 2020; Shadiev et al., 2020). These three studies focused on the development of AI-assisted systems for vocabulary learning. They each developed a system which allowed students to take pictures with smartphones and then upload them to the web, to be translated using Google Translate. Learners could make text or image annotations on the smartphone systems to support their vocabulary learning. However, the systems did not directly provide an object detection function or show how to spell or pronounce the word in addition to the translation function.

Shadiev et al. (2020) developed a learning system featuring image-to-text recognition (ITR) technology to support Russians learning English as a foreign language (see Figure 6). An Android-based mobile learning system was developed in their study and was installed on tablet PCs. In contrast to traditional image retrieval by typing in keywords, this system allowed users to search by submitting a sample image as their query. The Google Images service was employed for the ITR process. This service allows users to search the Web for image content. In comparison, the learners in the control group used the conventional approach, looking at the images that corresponded to the questions in their textbook. The experimental group outperformed the control group on the vocabulary tests.

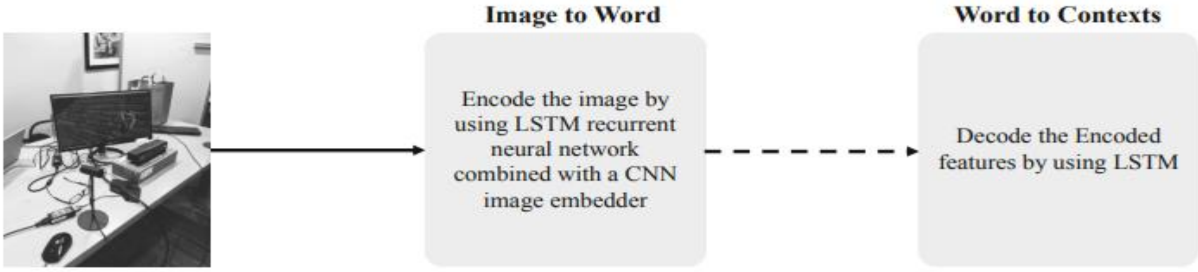
Figure 6. ITR system image-to-text recognition process (Shadiev et al., 2020)



(a) Take photo (b) Upload to the ITR system (c) Use the Google Image service to search the Web for image content

Hasnine et al. (2019) and Hasnine et al. (2020) created a system that allows students to take pictures that they upload as feedback (see Figure 7). The system then examines the visual content of these images and creates unique learning contexts on the basis of the visual content. Language learners could connect their previous knowledge with new knowledge by using this system, allowing them to review and recall the previously acquired vocabulary. Figure 7 illustrates the architecture of the model. In this model, the image encoder is a deep convolutional neural network (CNN), which is widely adopted for object recognition and detection tasks. The architecture of the model uses the Long Short-Term Memory (LSTM) network which is trained as a language model conditioned on image encoding. In the LSTM network, words in the captions are represented with an embedding model where each word is associated with a fixed-length vector representation that is learned during training.

Figure 7. Hasnine's et al. (2019; 2020) system image-to-word recognition process



3. Methodology

This study aimed to examine the effect of the AI-based ODT app on different ability EFL students' vocabulary learning using a quantitative approach. A randomized subjects, pre-test–post-test control group experimental design was adopted. The difference between the experimental group (EG) and the control group (CG) was the aid used to support task completion during the classroom activity sessions. Each of the learners in the EG was given a smartphone with the AI-based ODT app to finish the task guided by the instructor right after the teaching session, whereas the learners in the CG used Google Translate for the task; that is, they were given a vocabulary worksheet related to the teaching content. The dependent variable of this study was the different ability students' vocabulary performance on the post-test, and the independent variable was the different teaching platforms the students experienced (the AI-based ODT app vs. Google Translate).

Each group's average difference between the pre- and post-tests was calculated, and the average difference scores were compared to determine whether the experimental treatment produced a greater change than the control situation.

3.1. Participants

There were 72 participants involved in this research. They were students from a public elementary school located in a rural area in central Taiwan, and their native language was Mandarin Chinese. The average age of the participants was 11 years, with a range of 10 to 12 years. The participants had received formal English education in their elementary school for at least one year.

Each participant took the First Step level examination of Anglia English Speakers for Other Languages International Examinations (First Step Level of Anglia ESOL Exams) (Anglia Examinations England, 2014). The total score for the examination is 100 points. Those who passed the examination (with a passing score of 50) could understand the first useful structures such as classroom commands and how to introduce themselves, as well as approximately 100 words from familiar categories such as numbers, family members, colors, and household items. The mean score for the participants was 76.83.

The participants were divided into higher-level and lower-level groups based on the results of the scores of the First Step Level of Anglia ESOL Exams. Stratified sampling was then used to select samples from different level participants for the two groups—the experimental group and the control group—which ensured that the two groups were equivalent in terms of knowledge and ability.

3.2. Instruments

3.2.1 Vocabulary tests

In this research, a test sheet was designed to test the participants' vocabulary recognition. The 50 target words were selected from a daily market shopping experience, including 29 types of fruit and 21 vegetables. The participants were required to write the Chinese translation of the English words. The total score of the test was

100 points, with every question worth 2 points. The reliability coefficient value of the test was .93, indicating high reliability.

3.2.2. Class instruction and classroom activity

We planned four target vocabulary lessons, which followed the Presentation, Practice, Production model (PPP model) that many foreign language teaching course books are based on (Hutasoit et al., 2020). In the Presentation and Practice stages, both the CG and EG followed the same instruction. However, in the Production stage, the information exchange activity was applied to the CG by using a crossword puzzle, while in the EG, a role play activity was conducted.

Presentation (10 mins): In this stage, the instructor began the class lessons by presenting the target words and model sentences. One of the researchers as the instructor introduced the course first and then started to teach the vocabulary. Each week, the instructor began the presentation with 8-10 target words and the corresponding sentence structures. The instructor used flash cards to lead into the structure to be taught.

Practice and production (80 mins): In the practice stage, the new language was practiced by the students in a controlled manner. Students repeated the new words together and then separately right after the teacher to be able to say them correctly. In the production stage, the students used the language in context. For the EG, each student had a smartphone with the AI-based ODT app, which they could manipulate to scan the fruit or vegetables, listen to the pronunciation, and say the words correctly. The students in the CG were shown the same fruit or vegetables, and used Google Translate by typing the Chinese. They were asked to pronounce the word in English. A snapshot of the exploration time of the EG is shown in Figure 8.

Figure 8. Students used the AI-based ODT app to learn English vocabulary

3.2.3. The AI-based ODT app

The research team utilized TensorFlow to train the object detection model with the daily fruit data, and provided the corresponding English words for the pictures. After training, the model was exported to the mobile app as an English language learning tool. For training the object detection model, the camera was used to take 360° photos of each fruit, and then approximately 500 different angles of the fruit were chosen from the photos, which were arranged in a folder that was labeled with the name of the fruit (Chang et al., 2019; Liu et al., 2021).

Mobile Net was trained using TensorFlow's open source code tensorflow-for-poets-2, and the bottleneck model was used to implement the image recognition model. The training image was 224 pixels wide, and MobileNet's initial relative size was set to 1.0 (see Figure 9). Furthermore, Intel Core i5 2.70 GHz CPU, 4 GB RAM, and a 100-Mbps network environment were used in the research. The trained modules were connected to the mobile phone by using Android Studio 3.0. From the training dataset, two types of fruit, namely oranges and apples, were chosen for testing. The rationale for choosing these two types of fruit was that their appearance is similar (see Figure 9), making it easy to confuse the image recognition model.

We split the dataset into 10 parts and used a 10-fold cross-validation to determine the model's accuracy. Cross validation is a method applied to a model and a data set in an effort to estimate the out of sample errors. The relative sizes and pixel settings of different Mobile Nets were then recorded. Following the training, 16 models were obtained, and the data from TensorFlow were used for further analysis. We used a tensor board, which is a tool for providing the measurements and visualizations needed during the machine learning workflow, to compare the recognition rates of 16 image recognition models trained with different parameters, and then we chose the best image recognition model for use on mobile phones. Figure 10 shows the training results of the app accuracy. The results revealed that this study's image recognition model was very accurate (98%) (Chang et al., 2019; Liu, 2018). Figure 8 shows that the EG students used the app to identify the objects (e.g., cabbage and carrot) and learned English words with English/Chinese word translations and pronunciation.

Figure 9. Image recognition results for orange (93%) and apple (97%)

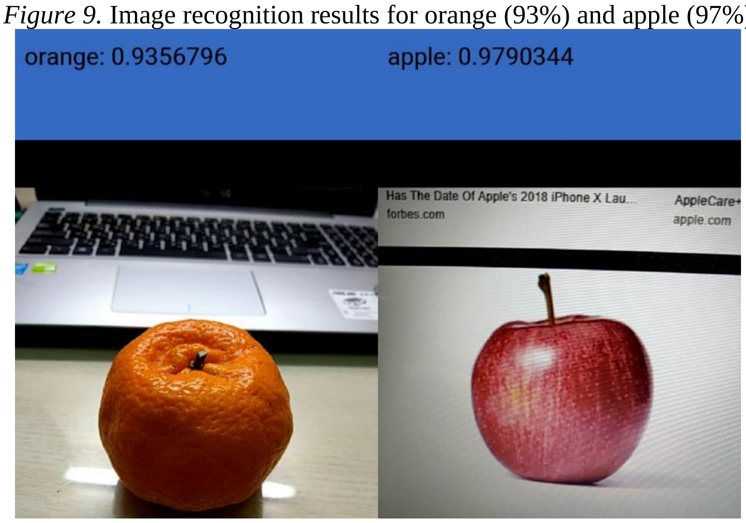
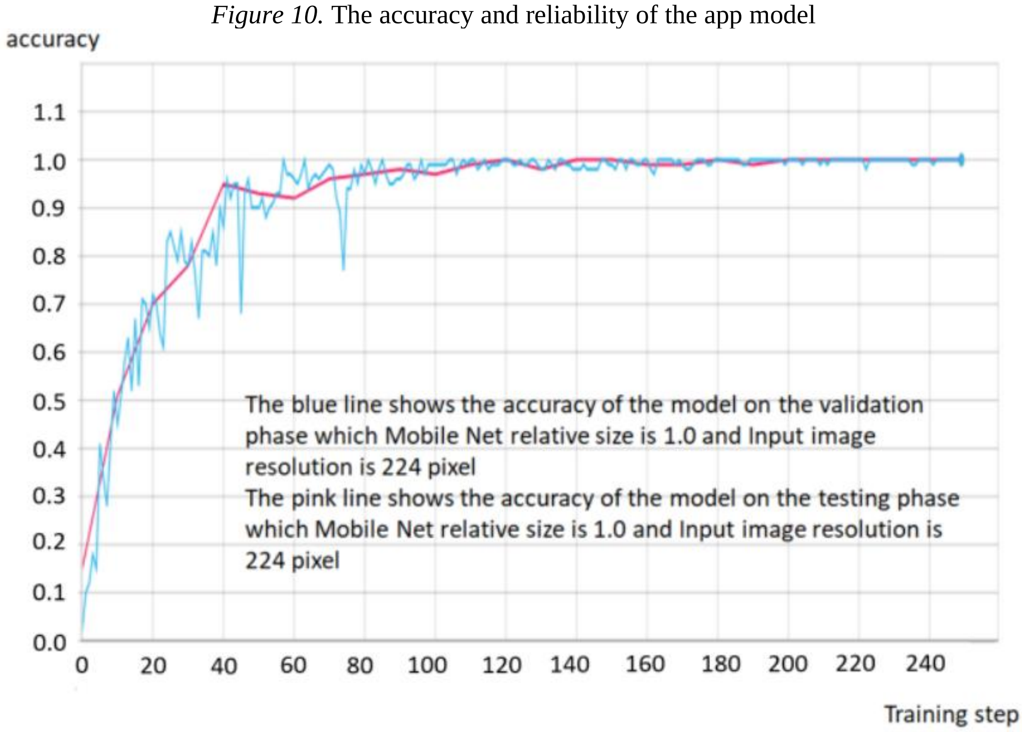


Figure 10. The accuracy and reliability of the app model



3.3. Procedure

The total duration of the experiment was 11 weeks. In the first week, all participants were asked to take the First Step Level of Anglia ESOL Exams to measure their English proficiency at the beginning of the experiment. Stratified sampling was then used to select samples from different level participants for the EG and CG. In the second week, the vocabulary pre-test was administered to represent the learners' prior knowledge of the English learning content. In the following paragraphs, the procedure and the content of six vocabulary lessons will be illustrated.

In order to reduce the practice effects, the researchers provided a two-week warm-up story reading exercise before the experiment began. From weeks 5 to 10, six vocabulary lessons were conducted, with one per week. There were 50 target words in total. The vocabulary related to fruit was taught in lessons 1 to 3, including star fruit, dragon fruit, and grapes. Next, the target vocabulary related to vegetables was taught in lessons 4 to 6, such as cauliflower, cucumber, and garlic. In each 90-min class, we followed the PPP model to teach the target vocabulary and provide opportunities for practice. During the class periods with the PPP model, all the participants learned the target vocabulary and applied it to the model sentences during the presentation stage. In the practice and production stages, the two groups were separated into two different classrooms with different activities for practice during which they applied the vocabulary learnt in class. For the EG, each student had a smartphone with the AI-based ODT app to scan the fruit or vegetables. The students in the CG were shown the same fruit and vegetables and used Google Translate by typing in the Chinese. Both groups were asked to listen to the pronunciation and to try to say the words correctly.

After six vocabulary lessons, in week 11, all participants were asked to take the vocabulary post-test. The post-test had different permutations of the questions from the pre-test. An independent t test was conducted using the vocabulary pre-test and post-test to compare the difference of the EG and CG data as the independent variable.

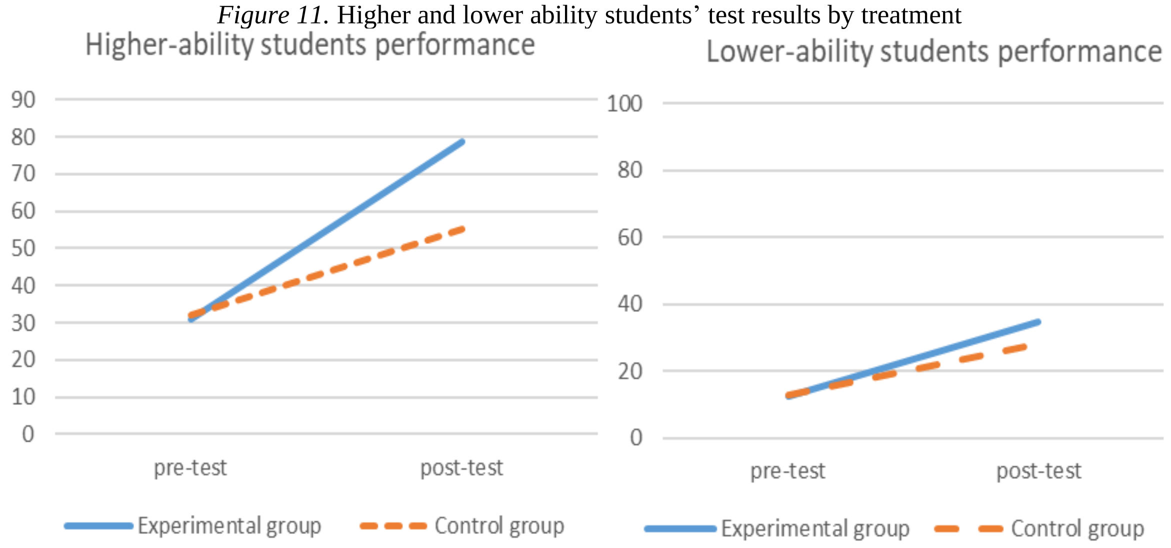
4. Results

The pre-test mean score of the EG group was 18.66 and that of the CG group was 23.66. The mean score of all participants was 21.17. This result indicated that there was no statistically significant difference between the EG and CG groups in the pre-test (t = 1.16, p = .25), indicating that these two groups had similar vocabulary ability prior to the experiment.

Table 2. Descriptive data of the different ability students' test results by treatment

	N	Pre-test		Post-test		*MD*
		M	*SD*	*M*	*SD*	
Higher-ability						
EG	18	30.89	10.91	78.67	16.92	50.00
CG	18	32.22	24.00	55.33	24.49	22.00
Total	36	31.56	17.96	67.00	23.88	36.00
Lower-ability						
EG	18	12.33	4.92	34.67	16.21	22.34
CG	18	13.00	10.23	28.00	3.50	15.00
Total	36	12.67	7.88	31.33	12.04	19.29

Figure 11. Higher and lower ability students' test results by treatment



We then employed the pre-test scores for further analysis. The students were divided into different ability groups according to their pre-test scores. Students who scored above the mean score in the pre-test were identified as higher-ability learners and those who scored at or below the mean score were identified as lower-ability learners. This resulted in 36 students in the app treatment (higher-ability M = 28.67, lower-ability M = 12.33) and 36 in the traditional treatment (higher-ability M = 33.33, lower-ability M = 13.00). There was also no significance difference between the higher-ability students (t = .77, p = .44) and lower-ability students (t = .51, p = .62) in the two groups on the pre-test (see Table 2, Figure 11).

The data analysis for student achievement was a 2 (Experimental treatment vs. Traditional treatment) × 2 (Ability Level: Higher-ability vs. Lower-ability) Two-Way Repeated measurement ANOVA on the pre-test and post-test (see Table 3)**.** The analysis included between-subjects variables such as treatment and skill level, as well as within-subjects variables such as test occasion and problem type. Table 2 shows the post-test results for the two types of treatment of the different ability students. The data on each variable's achievement is discussed below. Significant differences in English entering understanding and treatment were found using a 2 x 2 ANOVA. In terms of English entering knowledge level, students with higher-ability English entering knowledge outperformed those with lower-level English entering knowledge by a substantial margin (M = 67.00 and M = 31.33, respectively), $F(1, 72) = 78.93$, MS = 22898.00, $p = .00^*$. For treatments, the mean correct scores were 41.67 for the OTR translation app subjects and 56.67 for the traditional subjects, $F(1, 72) = 13.96$, MS = 4050.00, $p = .00^*$. The 2 x 2 ANOVA also yielded a significant two-way interaction for treatment by English ability, $F(1, 72) = 4.31$, MS = 1250.00, $p = .04^*$. This two-way interaction reflected the fact that the OTR translation app higher-ability students had considerably higher scores than the traditional higher-ability students (M = 78.67 vs. M = 55.33), but the lower-ability students of the two treatments had similar scores (M = 34.67 vs. M = 28.00) on the post-test.

Table 3. Test of between-subject effects

	SS	*df*	*MS*	*F*	*p*
Treatments	4050.00	1	4050.00	13.96	$.00^*$
Ability	22898.00	1	22898.00	78.93	$.00^*$
Treatments x Ability	1250.00	1	1250.00	4.31	$.04^*$
Error	19728.00	68	290.12		
Total	221976.00	72			

Note. $^*p < .05$.

As the follow-up to the two-way interaction, further univariate analysis of variance revealed that the slight difference in post-test scores by treatment for students with lower-ability was not statistically significant (F = 2.91, p = .54), whereas the difference in scores favoring the AI-based ODT app over the traditional treatment for students with higher-ability was significant (F = 11.07, $p = .00^*$).

5. Discussion

In terms of vocabulary attainment, there was a substantial difference between the experimental and control groups; students who learned using the AI-based ODT app technology scored higher on the post-test than those who learned using the non-AI technique. This finding is consistent with the findings of previous research (Asmana & Arifani, 2021; Hasnine et al., 2019; Hasnine et al., 2020; Katemba, 2021; Shadiev et al., 2020). First of all, the experimental learners learned vocabulary more actively than their counterparts. By integrating the word translation and pronunciation, this type of learning based on the Dual-Coding Theory (DCT) differs from rote learning (Chuo & Yen, 2014; Wu & Huang, 2017). It is easier for learners to remember vocabulary if both the visual and verbal channels are integrated simultaneously while learning the input information (Kassim, 2018).

According to DCT (Clark & Paivio, 1991; Kanellopoulou et al., 2019; Kassim, 2018), a cognition theory, a learner's memory consists of two separate but interrelated verbal and visual codes for processing information. Interestingly, there exists an interconnection between the two separate systems which facilitates dual coding of information if not activated independently. Psychologists have demonstrated how our minds respond well to words that form a picture, and some studies have found that individuals who read illustrated text outperformed those who read text alone (Mayer & Sims, 1994).

In this study, we examined the interconnection between the combination of the AI-based ODT app which is responsible for visual capability and speech pronunciation and which provides the verbal capability. The visualized spoken words, based on DCT, may not only improve a learner's recollection of English vocabulary, but also stimulate other information processing such as feature identification and linkage with past knowledge. The detailed examination of the pre-post test results provided an insight into how the combination, actualized through different teaching platforms (AI and Non-AI with object detection-enabled or disabled), scaffolds children's experiences and learning of a second language (Dalim et al., 2020).

In both experiments, the participants who used the AI-based ODT app platform had a significantly higher knowledge gain than those who used the non-AI platform (Google Translate app). Unlike the non-AI teaching platform which provides limited ability to manipulate the visual feedback, participants were able to scan objects with the AI-based ODT app in a more interactive way. The ability for participants to hold the mobile phone to scan and see the input on the screen while remaining in their real environment made the learning more interesting. This AI-based ODT app learning activity ensured active learning. Kim (2011) believed that if students are actively engaged in the cognitive processes of vocabulary acquisition, they will learn more words and retain them for longer. Moreover, because the experimental learning task was considerably more complicated and challenging than the traditional activities, the higher-ability students were more involved in this activity than the lower-ability students.

Another reason for the above-mentioned outcome was that the experimental learners' first encounter with this technology was a novel and exciting experience for them. According to previous research, when new technologies are used in education, they motivate students to participate in the learning process. As a result, students are more engaged in the learning process, making it easier for them to comprehend the material (Sahin & Yilmaz, 2020). Moreover, the AI-based ODT app is interesting for students and attracts their attention (Hasnine et al., 2019; Hasnine et al., 2020; Shadiev et al., 2020). It also helps them achieve their goals. According to the literature, AI may maximize children's early linguistic development, as well as increase attendance and provide students with personalized instruction (Porayska-Pomsta, 2016).

According to Venkatesh and Davis (2000), if learners find a system to be simple to use and useful, they will use it again. The findings revealed that the majority of the experimental group students approved of the learning system and were pleased with its use, as their interaction with the system was simple and clear. The acceptance of such systems is largely attributed to the accuracy of their recognition processes (Shadiev & Sun, 2020). To train the picture recognition model, a camera was used to take 360° photos of each fruit, and then approximately 500 different angles of the fruit were chosen from the photos. The results revealed that the proposed image recognition model was very accurate (98%) (Chang et al., 2019; Liu, 2018). As a result, the learners found the AI-based ODT app to be easy to use. The greatest challenge observed in the use of the Google Translate system was the difficulty the children faced in synchronizing their hand movements when typing the Chinese or English while searching for the translation; this was especially true for the lower-ability students.

Despite the positive results of this study, there are some limitations that should be noted. For example, for training the picture recognition model, we selected only 50 types of fruit and vegetable because of the limitations of availability and time. As a result, larger database studies are needed to add more categories to ensure the validity of the methodological approach. We also found that the AI based-ODT app benefited the higher-ability students more than the lower-ability students. It would be more beneficial if the design of the app could be modified to help the lower-ability students learn vocabulary more easily. This would provide a more significant contribution to the topic of language learning.

6. Conclusion and future work

In this article we have presented an AI based-ODT app for teaching young children who are non-native English speakers the English vocabulary for basic fruit and vegetables. Based on previous studies, our system is the first AI based-ODT app language-learning tool designed to teach young children. The objectives of this study were to ascertain how effective the AI based-ODT app combined with the DCT design was compared to the use of the non-AI Google Translate app, and to explore if the use of object detection translation was able to increase the effectiveness of vocabulary learning. The findings suggest that the AI-based ODT app could be useful as a teaching aid for young children, as it boosts learning engagement and knowledge gain. Real-time interaction encourages children to delve more deeply into the learning materials. The participants had positive experiences of interacting with the AI-based ODT app interface over the non-AR interface. Our findings conclude that the speech-enabled AR interface is usable for younger children even with little or no prior AR experience, and provides a motivating factor for their foreign language learning. The results show significant evidence of knowledge gain and a positive inclination to use the AR interface over the Google Translate interface.

In terms of future work, more research is needed to discover the potential of AI-based ODT for young children's activities. This study could be replicated in other domains where learners could learn about objects by differentiating their knowledge. For example, learners learning sign language for special purposes may use the app to take photos of sign language to learn the corresponding translations. We also suggest that future studies elaborate more on the benefits of ODT for language learning, and determine which of the three functions of the app (picture, text, and pronunciation) is the most beneficial and meaningful for language learning.

References

Almadhady, A. A., Salam, A. R. H., & Baharum, H. I. (2020). The Motivation of Arab EFL university students towards using MALL applications for speaking improvement. *Universal Journal of Educational Research*, *8*(11), 23-36.

Alnujaidi, S. (2021). Adoption of mobile assisted language learning (MALL) in Saudi Arabian EFL Classrooms. *Journal of Language Teaching and Research*, *12*(2), 312-323.

Asmana, C. H., & Arifani, Y. (2021). Mobile-assisted language learning based using scientific approach to improve students' vocabulary. *Journal of English Teaching, Literature, and Applied Linguistics, 4*(1), 43-51.

Anglia Examinations England (2014). *Handbook for teachers: Full examination syllabus and specifications for the revised exams 2014/2015*. Anglia Examinations Syndicate Limited.

Bax, S. (2003). CALL: Past, present and future. *System, 31*(1), 13-28.

Bax, S. (2011). Normalisation revisited: The Effective use of technology in language education. *International Journal of Computer-Assisted Language Learning and Teaching, 1*(2), 1-15.

Burston, J. (2014). MALL: The Pedagogical challenges. *Computer Assisted Language Learning, 27*(4), 344-357.

Chang, H. Y., Liu, P. L., Wang, W. C., & Chen, T. C. (2019). English learning tool for elders-Design of instant fruit identification system with TensorFlow. *Proceedings of* the *2019 IEEE International Conference on Consumer Electronics.* https://doi.org/10.1109/ICCE-TW46550.2019.8991988

Chang, M. M., & Hung, H. T. (2019). Effects of technology-enhanced language learning on second language acquisition: A Meta-analysis. *Journal of Educational Technology & Society*, *22*(4), 1–17.

Chai, C. S., Lin, P. Y., Jong, M. S. Y., Dai, Y., Chiu, T. K. F., & Qin, J. (2021). Perceptions of and behavioral intentions towards learning artificial intelligence in primary school students. *Educational Technology & Society, 24*(3), 89-101.

Chatterjee, S., & Bhattacharjee, K. K. (2020). Adoption of artificial intelligence in higher education: A Quantitative analysis using structural equation modeling. *Education and Information Technologies*, *25*(5), 3443-3463.

Chun, D. M. (2016). The Role of technology in SLA research. *Language Learning & Technology, 20*(2), 98-115.

Chuo, T. W. I., & Yen, S. C. H. (2014). The Learning journey of college at-risk EFL students in Taiwan: An exploratory study. *The Asian EFL Journal Quarterly, 16*(2), 44-68.

Clark, J. M., & Paivio, A. (1991). Dual coding theory and education. *Educational Psychology Review, 3*(3), 149-210. https://doi.org/10.1007/BF01320076

Copeland, M. (2016). What's the difference between artificial intelligence, machine learning and deep learning? https://blogs.nvidia.com/wp-content/uploads/2016/07/Deep_Learning_Icons_R5_PNG.jpg.png.webp

Dalim, C. S. C, Sunar, M. S., Dey, A., & Billinghurst, M. (2020). Using augmented reality with speech input for non-native children's language learning. *International Journal of Human-Computer Studies, 134*, 44-64. https://doi.org/10.1016/j.ijhcs.2019.10.002

de Carvalho, T. B., Sibaldo, M. A., Tsang, R., Cavalcanti, G. D., Sijbers, J., & Tsang, J. (2018). Intensity patches and pegion patches for image recognition. *Applied Soft Computing*, *62*(2018), 176-186.

Dreyer, C., & Nel, C. (2003). Teaching reading strategies and reading comprehension within a technology-enhanced learning environment. *System, 31*(3), 349-365.

Eslami, R., & Ahmadi, S. (2019). Investigating the role of educational media on secondary school students' learning process improvement in Jahrom city. *Journal of Humanities Insights*, *3*(1), 13-16.

Fu, C. (2016, May 12). *Google Translate real-time camera translation new function, full support for traditional Chinese recognition*. https://www.cool3c.com/article/106214

García Botero, G., Questier, F., & Zhu, C. (2019). Self-directed language learning in a mobile-assisted, out-of-class context: do students walk the talk? *Computer Assisted Language Learning*, *32*(1-2), 71-97.

Gharehblagh, N. M., & Nasri, N. (2020). Developing EFL elementary learners' writing skills through mobile-assisted language learning (MALL). *Teaching English with Technology*, *20*(1), 104-121.

Ghorbani, N., & Ebadi, S. (2020). Exploring learners' grammatical development in mobile assisted language learning. *Cogent Education*, *7*(1), 1704599. https://doi.org/10.1080/2331186X.2019.1704599

Giovany, S., Putra, A., Hariawan, A. S., & Wulandhari, L. A. (2017). Machine learning and sift approach for Indonesian food image recognition. *Procedia Computer Science*, *116,* 612-620.

Good, O. (2015, July 29). *How Google Translate squeezes deep learning onto a phone.* https://ai.googleblog.com/2015/07/how-google-translate-squeezes-deep.html

Goodfellow, I., Bengio, Y., Courville, A., & Bengio, Y. (2016). *Deep learning*. MIT press.

Gu, X. (2019, July 10). *Google Translate's instant Camera Translation gets an upgrade*. https://blog.google/products/translate/google-translates-instant-camera-translation-gets-upgrade/

Gulson, K. N., Murphie, A., Taylor, S., & Sellar, S. (2018). *Education, work and Australian society in an AI world: A Review of research literature and policy recommendations.* Gonski Institute for Education.

Hasnine, M. N., Flanagan, B., Akcapinar, G., Ogata, H., Mouri, K., & Uosaki, N. (2019). Vocabulary learning support system based on automatic image captioning technology. *Lecture Notes in Computer Science, 11587*, 346-358. doi.org/10.1007/978-3-030-21935-2_26

Hasnine, M. N., Mouri, K., Akcapinar, G. Ö. K. H. A. N., Ahmed, M. M. H., & Ueda, H. (2020). A New Technology Design for Personalized Incidental Vocabulary Learning using Lifelog Image Analysis. *In Proceedings of the 28th International Conference on Computers in Education (ICCE2020)* (pp. 516-521).

Hubbard, P. (2013). Making a case for learner training in technology enhanced language learning environments. *CALICO Journal, 30*(2), 163-178.

Hutasoit, E. V., Elfiza, R., & Subroto, G. (2020). The Effect of using PPP method on students' speaking skill. *Student Online Journal, 1*(2), 258-263.

Hwang, G. J. (2003). A Conceptual map model for developing intelligent tutoring systems. *Computers & Education*, *40*(3), 217-235.

Islam, A. B. M. S., & Hasan, M. (2020). The Effectiveness of mobile assisted language learning (MALL) on ESL listening skill. *NOBEL: Journal of Literature and Language Teaching*, *11*(2), 188-202. https://doi.org/10.15642/NOBEL.2020.11.2.188-202

Jin, L. (2018). Digital affordances on WeChat: Learning Chinese as a second language. *Computer Assisted Language Learning, 31,* 27-52.

Kaliisa, R., Palmer, E., & Miller, J. (2019). Mobile learning in higher education: A Comparative analysis of developed and developing country contexts. *British Journal of Educational Technology*, *50*(2), 546-561.

Kamasak, R., Özbilgin, M., Atay, D., & Kar, A. (2021). The Effectiveness of mobile-assisted language learning (MALL): A Review of the extant literature. In A. S. Moura, P. Reis, & M. N. D. S. Cordeiro (Eds.), *Handbook of Research on Determining the Reliability of Online Assessment and Distance Learning* (pp. 194-212). https://doi.org/10.4018/978-1-7998-4769-4

Kanellopoulou, C., Kermanidis, K. L., & Giannakoulopoulos, A. (2019). The Dual-coding and multimedia learning theories: Film subtitles as a vocabulary teaching tool. *Education Sciences, 9*(3), 210. https://doi.org/10.3390/educsci9030210

Kassim, W. (2018). Utilizing dual coding theory and animated images to enhance ESL students'vocabulary learning. *The English Teacher, 47*(3), 81-91.

Katemba, C. V. (2021). Enhancing vocabulary performance through mobile assisted language learning at a rural school in Indonesia. *Acuity: Journal of English Language Pedagogy, Literature and Culture, 6*(1), 1-11.

Kim, Y. (2011) The Role of task-induced involvement and learner proficiency in L2 vocabulary acquisition. *Language Learning, 61*(1), 100–140. https://doi.org/10.1111/j.1467-9922.2011.00644.x

Kong, F., He, H., Raynor, H. A., & Tan, J. (2015). DietCam: Multi-view regular shape food recognition with a camera phone. *Pervasive and Mobile Computing*, *19*, 108-121.

Krisbiantoro, B., & Pujiani, T. (2021). The Effectiveness of mobile-assisted language learning to teach writing. In L. Y. Rasnani, & L. Mareza (Eds.), *Empowering Human Development Through Science and Education* (pp. 1-15). https://proceedings.aecon.ump.ac.id/index.php/aecon/article/view/15/2

Lee, C. A., Tzeng, J. W., Huang, N. F., & Su, Y. S. (2021). Prediction of student performance in massive open online courses using deep learning system based on learning behaviors. *Educational Technology & Society, 24*(3), 130-146.

Lee, M. C., Chiu, S. Y., & Chang, J. W. (2017). A Deep convolutional neural network based Chinese menu recognition app. *Information Processing Letters, 128*, 14-20.

Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A. A., Ciompi, F., Ghafoorian, M., & Sánchez, C. I. (2017). A Survey on deep learning in medical image analysis. *Medical image analysis, 42*, 60-88.

Liu, P. L. (2016). Mobile English vocabulary learning based on concept-mapping strategy. *Language Learning & Technology, 20*(1), 128-140.

Liu, P. L. (2018). *Integrating concept map and AI image recognition translator on elders English vocabulary learning* (Report No. MOST 107-2635-H-415-002-). Ministry of Science and Technology.

Liu, P. L., Chen, C. J., Cheng, Y. H. (2021, June). *The Effect of image recognition mobile APP applied on EFL students' vocabulary learning* [Paper presentation]. 4th Pedagogy and Practice in Technology Enhanced Language Learning (PPTELL 2021), Taipei, Taiwan.

Liu, X., Liu, C. H., & Li, Y. (2020). The Effects of computer-assisted learning based on dual coding theory. *Symmetry, 12*(5), 701-713.

Loewen, S., Crowther, D., Isbell, D. R., Kim, K. M., Maloney, J., Miller, Z. F., & Rawal, H. (2019). Mobile-assisted language learning: A Duolingo case study. *ReCALL: The Journal of EUROCALL, 31*(3), 293-311.

Luckin, R., Holmes, W., Griffiths, M., & Forcier, L. B. (2016). *Intelligence unleashed: An Argument for AI in education.* Pearson Education

Lutfi, N. (2020). The Integration of MALL to enhance students speaking skill: An Autonomous learning model. *Journal of Foreign Language Teaching and Learning, 5*(1), 1-19.

Mayer, R. E., & Sims, V. K. (1994). For whom is a picture worth a thousand words? Extensions of a dual-coding theory of multimedia learning. *Journal of Education Psychology, 86*(3), 389-401.

Mezgec, S., Eftimov, T., Bucher, T., & Seljak, B. K. (2019). Mixed deep learning and natural language processing method for fake-food image recognition and standardization to help automated dietary assessment. *Public Health Nutrition, 22*(7), 1193-1202.

Nation, P. (2013). *What should every EFL teacher know?* Compass Publishing.

Oberg, A. (2011). Comparison of the effectiveness of a CALL-based approach and a card-based approach to vocabulary acquisition and retention. *CALICO Journal, 29*(1), 118-144.

Porayska-Pomsta, K. (2016). AI as a methodology for supporting educational praxis and teacher metacognition. *International Journal of Artificial Intelligence in Education, 26*(2), 679-700.

Rahimy, E. (2018). Deep learning applications in ophthalmology. *Current Opinion in Ophthalmology, 29*(3), 254-260.

Ramírez Verdugo, D., & Alonso Belmonte, I. (2007). Using digital stories to improve listening comprehension with Spanish young learners of English. *Language Learning & Technology, 11*(1), 87-101.

Ramos, R., & Dario, F. (2015). Incidental vocabulary learning in second language acquisition: A Literature review. *Profile Issues in Teachers' Professional Development, 17*(1), 157-166.

Ross, P. (1987). Intelligent tutoring systems. *Journal of Computer Assisted Learning, 3*(4), 194-203.

Sahin, D., & Yilmaz, R. M. (2020). The Effect of augmented reality technology on middle school students' achievements and attitudes towards science education. *Computers & Education, 144*(2020), 103710. https://doi.org/10.1016/j.compedu.2019.103710

Salloum, S. A., Alshurideh, M., Elnagar, A., & Shaalan, K. (2020). Machine learning and deep learning techniques for cybersecurity: A Review. In *The International Conference on Artificial Intelligence and Computer Vision* (pp. 50-57). Springer, Cham.

Shadiev, R., & Sun, A. (2020). Using texts generated by STR and CAT to facilitate student comprehension of lecture content in a foreign language. *Journal of Computing in Higher Education, 32*(2020), 561–581. https://doi.org/10.1007/s12528-019-09246-7

Shadiev, R., & Huang, Y. M (2020). Investigating student attention, meditation, cognitive load, and satisfaction during lectures in a foreign language supported by speech-enabled language translation. *Computer Assisted Language Learning, 33*(3), 301-326.

Shadiev, R., Sun, A., & Huang, Y. M. (2019). A Study of the facilitation of cross-cultural understanding and intercultural sensitivity using speech-enabled language translation technology. *British Journal Educational Technology, 50*, 1415-1433.

Shadiev, R., & Yang, M. (2020). Review of studies on technology-enhanced language learning and teaching. *Sustainability, 12*, 524. https://doi.org/10.3390/su12020524

Shadiev, R., Wu, T. T., & Huang, Y. M. (2020). Using image-to-text recognition technology to facilitate vocabulary acquisition in authentic contexts. *ReCALL 32*(2), 195-212. doi.org/10.1017/S0958344020000038

Shen, D., Wu, G., & Suk, H. I. (2017). Deep learning in medical image analysis. *Annual Review of Biomedical Engineering*, *19*(2017), 221-248.

Steiner, D. F., Chen, P. H. C., & Mermel, C. H. (2021). Closing the translation gap: AI applications in digital pathology. *Biochimica et Biophysica Acta: Reviews on Cancer, 1875*(1), 188452. https://doi.org/10.1016/j.bbcan.2020.188452

Tanaka, M. (2017) Examining EFL vocabulary learning motivation in a demotivating learning environment. *System, 65*(2017), 130–138.

Venkatesh, V., & Davis, F. D. (2000). A Theoretical extension of the technology acceptance model: Four longitudinal field studies. *Management Science, 46*(2), 186–204. https://doi.org/10.1287/mnsc.46.2.186.11926

Vincent, J. (2019, July 11). *Google's live camera translation is getting better AI and 60 new languages.* https://www.theverge.com/2019/7/10/20688682/google-translate-camera-instant-translation-function-languages-update-ai

Walker, A., & White, G. (2013). *Technology enhanced language learning: Connecting theory and practice.* Oxford University Press.

Weiland, C., & Yoshikawa, H. (2013). Impacts of a prekindergarten program on children's mathematics, language, literacy, executive function, and emotional skills. *Child Development*, *84*(6), 2112-2130.

Wu, T. T. (2014). The Use of a mobile assistant learning system for health education based on project-based learning. *Computers, Informatics, Nursing*, *32*(10), 497-503.

Wu, T. T. (2016). English reading e-book system integrating grouping and guided reading mechanisms based on the analysis of learning portfolios. *Journal of Internet Technology, 17*(2), 231–241.

Wu, T. T., & Huang, Y. M. (2017). A Mobile game-based English vocabulary practice system based on portfolio analysis. *Journal of Educational Technology & Society*, *20*(2), 265-277.

Xie, C., Tan, M., Gong, B., Wang, J., Yuille, A. L., & Le, Q. V. (2020). Adversarial examples improve image recognition. In L. O'Conner & H. Torres (eds.), *Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 819-828). The Institute of Electrical and Electronics Engineers. https://doi.org/10.1109/cvpr42600.2020.00090

Xie, H., Hwang, G. J., & Wong, T. L. (2021). From conventional AI to modern AI in education: Re-examining AI and analytic techniques for teaching and learning. *Educational Technology & Society, 24*(3), 85–88.

Xu, Q. (2020). Applying MALL to an EFL Listening and Speaking Course: An Action Research Approach. *Turkish Online Journal of Educational Technology-TOJET*, *19*(4), 24-34.

Yigitcanlar, T., Desouza, K. C., Butler, L., & Roozkhosh, F. (2020). Contributions and risks of artificial intelligence (AI) in building smarter cities: insights from a systematic review of the literature. *Energies, 13*(6) (2020), 1473-1511.

Yousef, M., Hussain, K. F., & Mohammed, U. S. (2020). Accurate, data-efficient, unconstrained text recognition with convolutional neural networks. *Pattern Recognition, 108*(2020),107482. doi.org/10.1016/j.patcog.2020.107482.

Zhang, C., & Lu, Y. (2021). Study on artificial intelligence: The State of the art and future prospects. *Journal of Industrial Information Integration, 23,* 100224. https://doi.org/10.1016/j.jii.2021.100224

Zhong, Y., Ma, A., Ong, S. Y., Zhu, Z., & Zhang, L. (2018). Computational intelligence in optical remote sensing image processing. *Applied Soft Computing*, *64*(2018), 75-93.

Jiang, S. (2023). Investigating Adolescents' Participation Trajectories in a Collaborative Multimodal Composing Learning Environment. *Educational Technology & Society*, *26*(3), 21-36. https://doi.org/10.30191/ETS.202307_26(3).0003

Investigating Adolescents' Participation Trajectories in a Collaborative Multimodal Composing Learning Environment

Shiyan Jiang
North Carolina State University, United States // sjiang24@ncsu.edu

ABSTRACT: The field of STEM education calls for a nuanced understanding of participation as participation measured by attendance provides limited information about student learning. This multiple case study contributes to a nuanced understanding of youth's participation trajectories in a multimodal composition project. In the project, fifth to eighth grade students worked in small groups to create multimodal science fiction stories in which they needed to propose creative solutions to issues related to climate change. In this study, I adopted two theoretical perspectives, disciplinary identity development and community of practice, to analyze participation trajectories with multiple sources of data. This study shows that STEM practices mediated by multiple modes can not only offer students flexibility in moving across forms of participation, but also open space for them to demonstrate their expertise as knowledge producers. Furthermore, this study suggests that the following strategies could be effective for broadening participation in STEM practices: supporting the development of reflective understanding of connections between disciplines through digital literacies, providing exposure in composing with multiple modes, focusing on building a close relation between self and digital artifacts, and offering flexibility in moving across interactional spaces. These insights shed light on broadening participation in other multimodal learning settings.

Keywords: Multimodal composition, Participation trajectory, Integrated STEM, Digital literacy, Broaden participation

1. Introduction

Integrating science, technology, engineering, and mathematics (STEM) and digital literacy practices holds the promise of broadening access to STEM (Hall & Coyne, 2005). Recent educational reforms have emphasized the importance of broadening adolescents' participation in and access to high-quality STEM practices (NGSS Lead States, 2013; NRC, 2017). The goal is to improve adolescents', especially those who are underserved or underrepresented, STEM competence so that they are prepared to succeed in college and future careers. Multimodal composition, a digital literacy practice in which people use different modes (e.g., text and visual) to represent ideas (Kress, 2003; Kress & Van Leeuwen, 2001), is engaging for adolescents. Integrating multimodal composition in STEM learning is particularly beneficial for minority students (Smith, 2014).

As the current generation of students grows up with increasing experience in consuming and creating digital products, STEM education needs to embrace this experience to enrich and support student learning. Adolescents are used to expressing themselves and connecting with others through multimodal composing. They often create and share digital multimodal artifacts, such as YouTube videos, personal blogs, and video games world widely (Smith, 2014). Students who were disengaged from school could be popular composers out-of-school. Although adolescents' out-school multimodal composing interests and expertise are gradually integrated into STEM curriculum (Smith & Dalton, 2016), much more needs to be done to fully connect multimodal composing practices and STEM practices to increase and broaden participation.

Meanwhile, researchers point out a lack of theoretical understanding and operationalized definition of participation in STEM education (Hrastinski, 2008; Malinen, 2015). Measuring participation from theoretical perspectives can guide us to better understand student learning as participation measured by attendance provides limited information about student learning. Youth tended to lose interest in certain disciplines gradually, especially STEM disciplines. Based on understanding the nature of participation in a nuanced way, we can develop design principles for broadening participation. This study will address this gap by closely examining adolescents' participation trajectories in a collaborative multimodal composing learning environment, drawing on an integrated theoretical framework (as described in Section 3).

2. Literature review

The literature shows that conceptualizing participation as a process of STEM identity development holds promises to provide novel and nuanced insights into student learning (Dou & Cian, 2022; Pinkard et al., 2017).

For example, Van Horne and Bell (2017) found that in the process of becoming core members in a community, one disciplinary identity (e.g., science identity) might be stabilized while the other might be destabilized. Furthermore, they showed that the alignment between students' future possible self and science practices could contribute to the stabilization of STEM identities. As another example, Nasir and Hand (2008) posited that participation in social and cultural practices was fundamentally related to practice-linked identities. They defined practice-linked identities as "a sense of connection between the self and the practice" (p. 147). This study demonstrated that an individual was more likely to participate actively when the person could connect self and STEM practices. Collectively, these studies highlighted the importance of understanding learning from the angle of participation in practices and identity construction.

However, much more effort is still needed in building a solid theoretical understanding of participation and more importantly, turning theoretical conceptualization into empirical operationalization of participation (Cohen et al., 2021; Hrastinski, 2008; Malinen, 2015). For instance, in Franz-Odendaal and colleagues' (2016) study, participation was measured at four levels: no STEM activity, low level, moderate level, and high level. This study elucidated that participation in STEM practices influenced students' likelihood of pursuing STEM careers and suggested that more sophisticated measures of participation were needed to unfold in what ways participation affected STEM career aspirations. Similarly, Boyce and colleagues (2014) found that active participants reported high motivation in pursuing STEM careers. The authors assessed participation based on how students navigated three interactional spaces: interaction with peers, technology, and the physical environment. This study called for applying a theoretical lens to operationalize participation for understanding the dynamics of participation across spaces. Consequently, there is a need of developing a complex and detailed understanding of students' learning experiences through conceptualizing and operationalizing participation from theoretical perspectives, which is the focus of this study.

Furthermore, researchers suggest that multimodal activities are engaging for adolescents to learn STEM practices (Jiang et al., 2021; Smith et al., 2021) and often open up opportunities for students to draw upon their cultural lifeworlds and out-of-school interests in empowering ways (Hull & Katz, 2006). Strides have been made in understanding student learning in activities that integrate multimodal composition and STEM; however, findings in this area tend to focus on reporting affordances of multiple modes in disciplinary learning (Krajcik & Sutherland, 2010). Meanwhile, the literature suggested that just examining one learning space was not sufficient to gain a holistic view of student learning. It's critical to gain a multidimensional understanding of students' multimodal products, composing processes, and reflections on their choices of modes in representing STEM knowledge and practices (Smith & Shen, 2017). This approach affords analyzing student learning across multiple dimensions, but it requires careful consideration of different data sources, especially in learning environments that offer students the flexibility of choosing modes of interest. This study addresses this research gap by presenting a multidimensional understanding of student learning in a multimodal composing environment, focusing on the construct of participation.

3. Theoretical framework

To examine participation trajectories, this study draws on an integrated theoretical framework. Specifically, this study frames students' participation from a community of practice perspective (Lave & Wenger, 1991) and a disciplinary identity development perspective (Van Horne & Bell, 2017). I conceive that students' participation can take various disciplinary *forms* and reach multiple *degrees* within face-to-face and online communities of practice.

This study frames student participation from a community of practice perspective (Lave & Wenger, 1991; Wenger, 1998). Lave and Wenger (1991) proposed that a community of practice consists of a group of people who interact with each other in the pursuit of a common goal. Over time, members of a community learn the appropriate work behaviors and norms as they increasingly participate in consuming and producing ideas. Participation in communities of practice can reach multiple *degrees*, including *breadth* and *depth* of participation (Smith & Shen, 2017). Breadth of participation was usually indicated by the amount of participation, such as time on task (Denault & Poulin, 2009; Handley et al., 2006; Preece, 2001; Zheng & Warschauer, 2015). Lave and Wenger (1991) described a range of depth of participation within communities of practice, from legitimate peripheral participation of newcomers to full participation of experts. According to Engle and Conant (2002), the moving from peripheral to full participation can be marked by levels of authorship of ideas. Engle and Conant proposed the significance of encouraging "students to be authors and producers of knowledge, with ownership over it, rather than mere consumers of it" (p. 404). Learners must aspire to become contributors and not simply consumers of knowledge produced by the community.

Furthermore, this study conceptualizes student participation as a process of disciplinary identity development. Students author various disciplinary identities prior to participating in any learning environment. As Van Horne and Bell (2017) explained, “all designed learning experiences involve a framing and cultivation of disciplinary identities” (p. 439). They also indicated that disciplinary identity development was a process of moving from legitimate peripheral participation to full participation in communities of practice through participating in domain-linked practices. In this perspective, a disciplinary identity (e.g., writing identity), is enacted when a person shows positive attitudes toward disciplines and interests in discipline-related careers (Archer et al., 2010), demonstrates normative disciplinary knowledge and practices (Carlone & Johnson, 2007), and gains recognition as a legitimate participant by self and others in communities (Lave & Wenger, 1991). Drawing on the construct of disciplinary identity, students could participate in communities of practice through taking various *forms* of participation. These forms are represented in the disciplinary roles that students play in a community in this study.

A community of practice lens examines the degree of participation while disciplinary identity development offers a way for investigating the form of participation. Combining the perspective of disciplinary identities with community of practice, we will gain insights into the degree of participation (including both breadth and depth) while students participate in STEM practices through different forms. Based on an integrated framework of participation, this study addresses the following research questions (RQ):

- RQ1: What are the trajectories of students’ *forms* of participation in a collaborative multimodal composing learning environment?
- RQ2: What are the trajectories of students’ *degrees (including both breadth and depth)* of participation in a collaborative multimodal composing learning environment?

4. Method

4.1. The STEM+L project

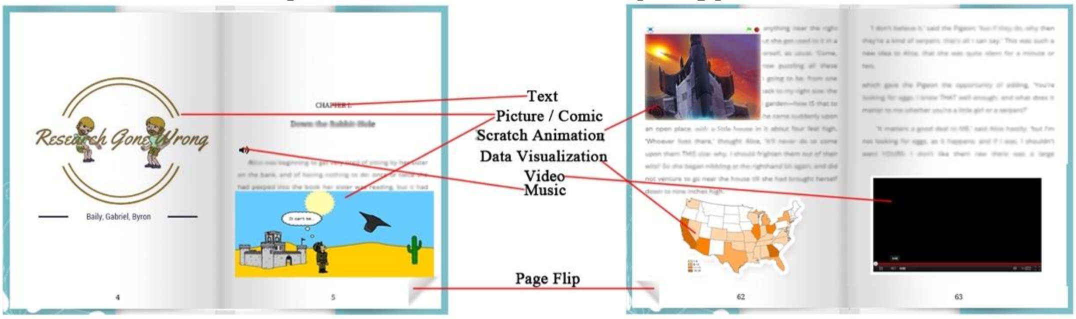
I report on a single iteration of a design-based research program (Cobb et al., 2003) that sought to broaden adolescents’ participation in integrated STEM+L (STEM and digital Literacies) practices. The project was hosted at a university in a large southeastern city in the United States. In the project, students worked collaboratively in small groups to create multimodal science fictions. In their science fictions, students were required to propose a creative solution to issues related to climate change.

Specifically, the project included three major phases: (1) one-week summer camp; (2) fall semester extension that incorporated online learning and six physical sessions on Saturdays during the fall; (3) a final show event to authentic audiences at an international science fiction film festival. During the summer, students attended the program every day from 9:30 am to 3:30 pm for a week. Throughout the week, students learned multimodal composing tools, were introduced to story writing and climate change, and visited science labs at the University. Furthermore, they worked in small groups of three to four to complete the first version of multimodal science fictions. In the fall extension, students continued working on the multimodal science fictions in small groups. At the end of the project, they presented multimodal science fictions to authentic audiences.

Role-taking. In small groups, each student selected one of the three roles: designer, scientist, and writer. Designers were responsible for creating visual and audio representations (e.g., comic); scientists were in charge of verifying and incorporating scientific information; writers were accountable for developing and writing the fiction plot. Each group should have at least one scientist and one writer. In addition, students could propose other roles, take hybrid roles (e.g., taking the role of scientist as a major role with a minor role in design), or change roles throughout the project. Despite differentiated roles, group members collaborated with each other on their individual and collective tasks.

Multimodal composition. Students learned multimodal composing tools, including Pixton (character and comic design application), Scratch (animation and game design program), Moviemaker (video editing program), and Pixlr (image editing application). Both Pixton and Scratch provide the functions of sharing their own artifacts, remixing others’ artifacts, and posting comments in the corresponding community. All groups used iKOS (Jiang et al., 2021) to compile multimodal artifacts into interactive flipbooks (Figure 1). iKOS is a web-based knowledge organization platform for individuals and/or groups to construct, share, and organize knowledge in multimodal representations. Users can embed artifacts created in the system (e.g., written narrative) and external artifacts (e.g., comics from Pixton) to generate an interactive flipbook. Despite these functions, students can interact with each other through rating, commenting, and co-editing entries within the system. The system also creates log data of students’ frequencies, durations, and sequences of computer actions (e.g., the timestamp when students click one entry). Beyond the presented technologies, students were free to bring in multimodal tools that they were familiar with, and the research team provided corresponding technology support.

Figure 1. iKOS, a multimodal composing platform



4.2. Participants

This study included a total of 42 students (19 Latinx, 14 Black, 4 White, and 5 other ethnic groups; 16 female and 26 male). These students were divided into two cohorts (cohort A, 22 students; cohort B, 20 students) due to resource limitations. Students were from public schools in the Southeastern U.S. area, with six 5th graders, fifteen 6th graders, eight 7th grades, and thirteen 8th graders. Four focal students in four different groups were selected for in-depth analysis of their participation trajectories: Olivia (all names were pseudonyms; seventh grader), Nick (sixth grader), Steve (sixth grader), and Saanvi (fifth grader). These four cases were selected for several considerations. First, they attended most or all the sessions so that a relatively complete profile for each of them could be captured. Second, these cases instantiate maximum variation (Flyvbjerg, 2006) in terms of gender, race, grade, and engagement in STEM+L practices throughout the program. The selection was also what Flyvbjerg (2006) called an informed-oriented selection: From field notes and review of their multimodal artifacts, I expected these cases to contain rich examples of scaling and assembling comparisons in participation trajectories.

4.3. Researcher’s participant observer role

My role in this study was a participant observer (Spradley, 1980) who was highly involved in designing learning activities, instructing and interacting with students, and collecting data. While interacting with students, I paid close attention to their role taking and multimodal composing processes and provided feedback when needed. As a participant observer, reflexive awareness of my own biases and positionality was crucial for establishing trustworthiness (Miles & Huberman, 1994). In order to establish trustworthiness in this qualitative study, I followed the following widely regarded standards: credibility, transferability, dependability, and confirmability (Erlandson et al., 1993).

4.4. Data collection

A variety of data were collected: pre-, mid-, and post-surveys for understanding students’ experiences of the project; student-generated multimodal artifacts; iKOS logging data; comments and ratings in Pixton, Scratch, and iKOS; video records of students’ physical participation together with the audio records of students’ conversations; group-based semi-structured interviews (Patton, 1990) for identifying students’ perceptions of teamwork, role taking, multimodal design decisions, and science learning; video reflections (i.e., students created videos to reflect on learning experiences); field notes; and physical materials (e.g., poster). Specifically, the research team used Qualtrics to administer online surveys. We conducted surveys three times: one on the first day of the summer camp, one on the fifth day of the summer camp, and one at the fifth session of the fall semester extension. In addition, in the fifth session during the fall, we conducted interviews with two groups and video and audio recorded the interviews.

Specific data sources for the four focal cases varied based upon students’ choices of tools, the availability of video and interview data, and relevance of data to each case. Olivia mainly worked in iKOS; Nick contributed mostly in designing comics in Pixton and editing entries in iKOS; Steve mainly designed comics and posted comments in Pixton; Saanvi spread her work in all composing platforms. Although Olivia worked in iKOS, we did not analyze her iKOS logging data because she typed in word documents and then pasted texts in iKOS. We also did not analyze Saanvi's logging data because she worked in her group member's account. Due to the nature of afterschool programs, we could not collect video recordings for all physical sessions of any focal students. However, we analyzed other sources of data to get a full picture of students' participation.

4.5. Data analysis

The data analysis was mainly qualitative in nature, informed by a grounded theory approach (Strauss & Corbin, 1998) to address research questions centered around students' forms and degrees of participation in STEM+L practices. Table 1 provides an overview of the data analytic process. First, surveys were qualitatively analyzed, and initial case summaries were created to show students' changes in the form of participation through role taking. Specifically, I examined students' responses to pre-, mid-, and post-surveys on attitudes toward disciplinary roles (i.e., designer, scientist, and writer) to show how forms of participation changed over time. I also connected their attitudes toward disciplinary roles with interests in disciplinary practices (e.g., design, science, and writing) and STEM careers to understand how forms of participation changed. Second, initial case summaries were created to show students' changes in the breadth of participation using frequencies of multimodal artifacts, edits in iKOS, and comments and ratings in multimodal composing platforms. Third, I conducted content analysis of multimodal artifacts and online interactions to examine the forms and depth of participation. Fourth, I used interaction analysis (Jordan & Henderson, 1995) to analyze the video recordings. Initially, I focused on "hot spots" (Jordan & Henderson, 1995) of interaction as they pertain to students' forms (e.g., designing comics to illustrate story plots as a designer) and degrees of participation (e.g., proposing new ideas). In a more systematic, second pass through the recordings, content logs and memos were developed to describe students' participation trajectories. The interaction analysis provides a fine-grained understanding of the form and degree of participation.

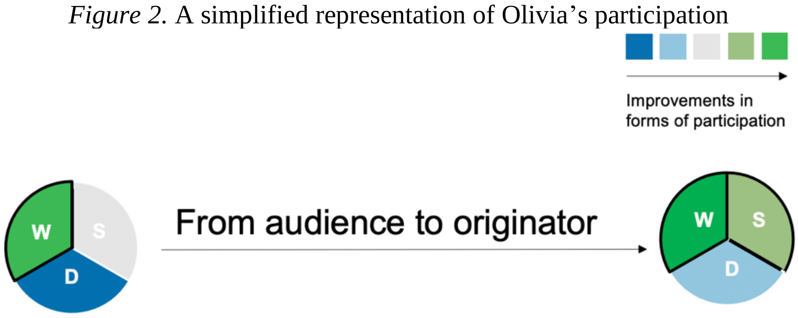
After analyzing these data sources, I discussed and revised case summaries with the larger research team in weekly meetings. In addition, I openly coded student reflections, video interviews, field notes, and physical materials that were connected to the form and degree of participation previously analyzed. In this process, I looked for connections to other data sources and new insights provided by students' perspectives and classroom observations. Afterward, I engaged in another round of revising cases with the larger research team. In this round of case revision, we followed Calabrese Barton and colleagues' (2013) methodology to craft participation trajectories based on focal events.

As an example, Figure 2 is a simplified representation of Olivia's participation. We first identified five focal events where Olivia changed her form or degree of participation through a systematic content analysis of case summaries. The focal events were creating animations, seeing a good example of enriching story with science, integrating space science into a new story, providing design ideas for textual narratives, and worrying about the design component of the story (Figure 3, a complete representation of her participation trajectory). After identifying focal events, we crafted her participation trajectory. Specifically, we first examined her changes in participation in disciplinary learning, including writing, science, and design. The color of S changed from grey to green. It means that Olivia's attitude toward science changed from neutral to positive based on survey data (Figure 2). This was a change in the form of participation. The border of pie changed from W to W and S. It means that Olivia recognized herself as a designer at the beginning and recognized herself as a writer and scientist at the end based on survey data. This was also a change in the form of participation. Olivia contributed to her group more as an audience when writing the first story, but shifted toward an originator when writing the second story based on content analysis of multimodal artifacts and interaction analysis of video recordings. The shape of the line (Figure 3) captured our qualitative interpretation of her change in originating ideas. The five focal events drove our qualitative interpretation of her changes in the depth of participation. To show the breadth of participation, we compared the number of words that Olivia wrote in two stories as she mainly engaged in this practice. Since she contributed a significant number of words in both stories, we conceptualize that her breath of participation did not change (i.e., the thickness of the line stays the same in Figure 3). Overall, driven by focal events, we crafted trajectories by first showing changes in forms of participation based on survey data and then conducting an in-depth qualitative analysis of multimodal artifacts and physical and/or online interactions to show degrees of participation.

Table 1. Overview of data analytic process (Strauss & Corbin, 1998)

Phase of data analytics	Data source						RQ1	RQ2
	Survey	Multimodal artifacts	Logging data	Comments and ratings	Video recordings	Others (e.g., interviews)	Forms of participation	Degrees of participation
1. Creating initial case summaries for students' forms of participation in terms of role taking	•						•	
2. Creating initial case summaries for students' breadth of participation in terms of online activities		•	•	•				•
3. Content analysis of multimodal artifacts and online interactions		•	•	•			•	•
4. Interaction analysis of video recordings					•		•	•
5. Revising case summaries	•	•	•	•	•		•	•
6. Open coding of students' perspectives on their participation and classroom observations						•	•	•
7. Revising case summaries	•	•	•	•	•	•	•	•

Figure 2. A simplified representation of Olivia's participation



Note. Each pie represents forms of participation through taking roles: W = Writer; S = Scientist, and D = Designer. Highlighted edges on pies indicate students' main roles. Colors of W/S/D represent strongly agree, disagree, neutral, agree, and strongly agree in having an interest in Writing/Science/Design.

5. Results

This section presents four cases to illuminate participation trajectories in the STEM+L project. Each case starts with a summary of the student's general background information and the group's composition. Then, I report overarching themes organized around forms and degrees of participation as framed in the two research questions.

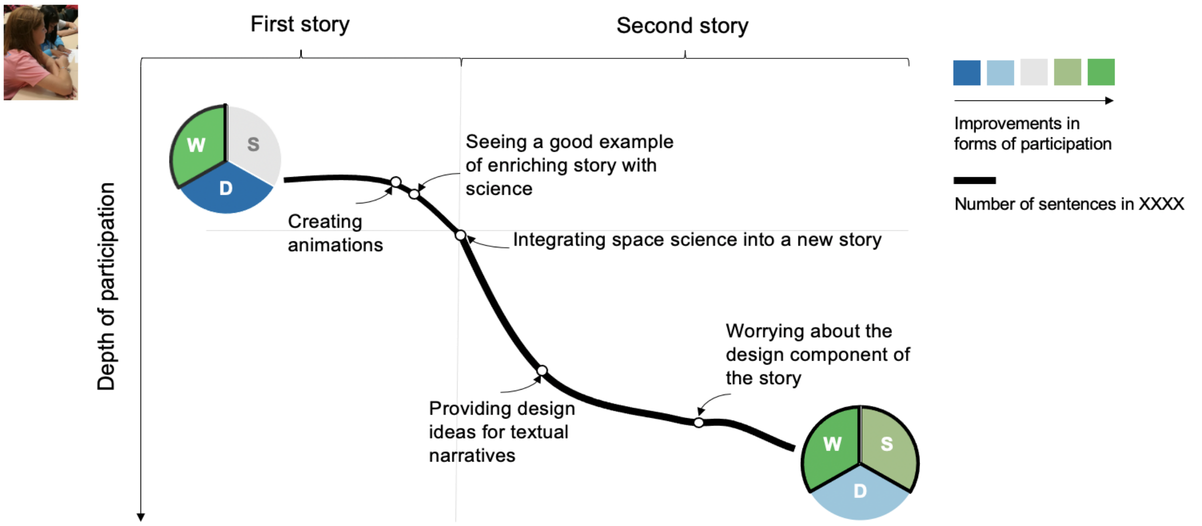
5.1. The case of Olivia: A trajectory from strong writer to emerging scientist

Olivia, a white female, was passionate about writing, selected the role of writer, and demonstrated herself as a strong writer throughout the program. In the summer, she joined a group of all seventh-grade students including Alyssa (African American female) who was the designer and Diego (Latino) and Bing (Asian male) who were both scientists. Diego led the development of the first story which was about a man who woke up after two years of a coma only to find out that the Earth was in darkness due to a long solar eclipse. Although Olivia was not excited about the story proposed by Diego, as the writer she wrote four chapters (1415 words in total). While working on the first story, she wrote textual narratives to match ideas in Diego's comics.

In the fall, because Diego and Bing didn't return to the program, Olivia and Alyssa decided to write a new story in spite of the instructors' advice to expand or revise the one the group had written in the summer. Inspired by a NASA news release about the discovery of a potential ninth planet, the two girls composed a fiction about a female middle schooler who wrote a story to describe her adventure on a new planet. Compared to the summer, Olivia was more active in developing the new story and provided Alyssa with design ideas. The pair ended up creating a story with three chapters (1787 words in total), two images, and four animations.

Figure 3 summarizes Olivia's participation in STEM+L practices. In this case, forms of participation changed mainly in the role of scientist while degrees of participation increased slightly in the breadth and greatly in the depth of participation.

Figure 3. Olivia's participation trajectories



Note. Each pie represents forms of participation through taking roles: W = Writer; S = Scientist, and D = Designer. Highlighted edges on pies indicate students' main roles. Thicker lines represent more breadth of participation.

5.1.1. Forms and degrees of participation

Olivia's case reveals a significant change in the form of participation, from taking the role of writer to taking the hybrid role of writer and scientist (Figure 3). Initially, Olivia only wanted to take the role of writer because of her passion and expertise in writing. She explained, "I love to write. I find myself transported to another world when getting my thoughts typed on a screen. My hands do the typing, and my brain the storytelling" (pre-survey). Understanding the connection between writing and science and learning examples of the integration between the two practices in animations helped Olivia expand from her comfort zone of writing into science learning. She successfully connected both and "worked well as writer and scientist" (post-survey).

As made evident in this case, the depth of participation changed significantly while the breadth of participation was pretty stable. Although having created many textural narratives in both fictions, Olivia contributed substantially in generating new ideas while composing the second fiction. The second fiction was inspired by a NASA news release about the discovery of a potential ninth planet. When Olivia came to a physical session, she discussed the news with one researcher and shared with the researcher the idea of creating a new story based on the ninth planet (Excerpt 1).

<u>Excerpt 1 [Group working on the second fall session]</u>
1 Olivia (*showing her screen*): Look at this. I do not know whether it's okay. Here, I wrote about the ninth planet, like how it looks and how the life there might be. I want to write a new story about the ninth planet.
2 Researcher: That's a lot. I would hesitate to change the whole story. But go ahead and jump in some ideas.
(*Olivia read her writing to the researcher*)
3 Olivia (*in an excited tone*): I am thinking that the ninth planet might like the surface of the Earth. There could be life on it, similar or might be different from us. Our story could show the possible things that are going on in the ninth planet.

The excerpt above demonstrated that the second story was driven by a science topic, the ninth planet, that Olivia was interested in. Olivia led the second fiction while two male students led the first one. While sometimes she expressed concern about missing comics in the story as the designer was only interested in creating Scratch animations, overall, her depth of participation improved over time. Being able to control the story development gave her a sense of ownership. It indicated that the cultivation of a sense of ownership would help increase the degree (especially depth) of participation.

In this case, the change from being a writer to a writer and scientist motivated Olivia to originate more science ideas in story writing. This suggests that the recognition of taking the role of scientist, impacted by changes in group dynamics (missing two scientists), contributed to improvements in both the breadth and depth of participation in science. Meanwhile, extended depth of participation in science, mediated by a sense of ownership, enabled her to originate most ideas in story writing and developed a positive attitude towards the role of scientist. In summary, Olivia's case illustrates that understanding the connection between disciplinary practices could result in changes in the form of participation while a sense of ownership is very important for deepening the depth of participation.

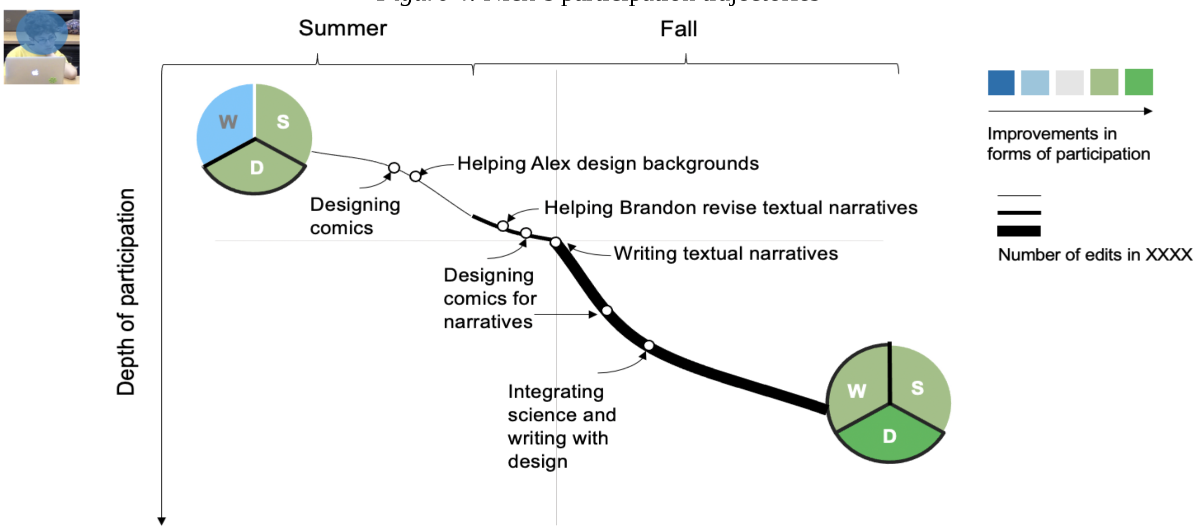
5.2. The case of Nick: A trajectory from diligent designer to motivated writer

Nick (sixth grade; Latino) took the role of designer, working with Alex (sixth grade; White male), the scientist, and Brandon (sixth grade; White male), the writer. The team composed a multimodal science fiction titled "What Would Happen if the World Stopped Spinning". In the fiction, three survivors, each representing a group member, discovered a new civilization after an asteroid struck the Earth and stopped the Earth from spinning. The team's multimodal composing processes involved the design of comics, photos, and images (Nick's major contribution), the development of textual narratives (Brandon and Nick's major contribution), and the creation of animations (Alex's major contribution).

To fulfill the role of designer, Nick mainly worked on creating comics during the summer camp and early fall, but engaged in designing visuals and writing textual narratives towards the end of the project. Overall, he contributed in creating a book cover and a photo, designing comics in Pixton, and writing story narratives in iKOS. Meanwhile, Nick was a collaborator and was the only one in the team who preferred working with others.

As shown in Figure 4, Nick extended his role of designer to the role of designer and writer. In this case, forms of participation changed in all three aspects (i.e., attitudes and interests, knowledge and practices, and recognition) while degrees of participation increased greatly in both the breadth and depth of participation.

Figure 4. Nick's participation trajectories

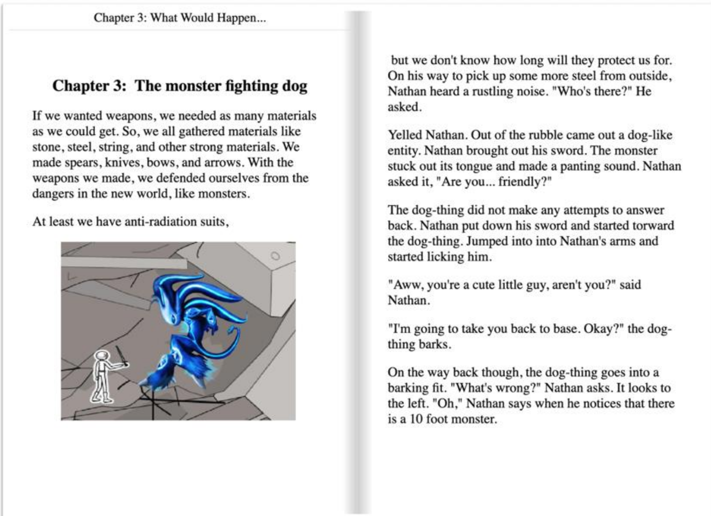


5.2.1. Forms and degrees of participation

Taking the role of designer as his primary form of participation, Nick's interest in being a designer stayed positive. He focused on designing comics for his team's work. Although the number of comics didn't change significantly over time, the quality did. This was reflected in his improved portrayal of nuanced actions of characters in his comics and better integration between comics and narratives, as well as incorporation of science concepts occasionally.

More interestingly, his attitude toward the role of writer turned from negative to positive (Figure 4). First, his preference in working with and helping others enabled him to enter writing practice. He started from reading textual narratives, then revised Brandon's texts, and finally wrote the story himself. This was consistent with his view of role taking as a way to try out different practices, including the writing practice that he had little confidence in, which in turn reinforced his self-recognition of being a writer. Second, the integration between design (i.e., creating comics) and writing (i.e., producing textual narratives) enabled Nick to see relevance and develop confidence in writing. As he was confident in design but not in writing in the beginning, extending that comfort zone to connect different disciplinary practices helped him build confidence in new areas. Third, the change from being a contributor to an originator in writing gave him more authority and agency in being a writer. While editing texts as a writer, besides revising existing texts as a more peripheral participant, over time, Nick added original story narratives as a more central participant. Lastly, infusing himself in his own character in the story (Figure 5) and being able to make himself visible in the team artifact might also contribute to a more positive attitude towards taking the role of writer.

Figure 5. Nick wrote a chapter about the character, Nathan, representing himself in the story


Chapter 3: What Would Happen...

Chapter 3: The monster fighting dog

If we wanted weapons, we needed as many materials as we could get. So, we all gathered materials like stone, steel, string, and other strong materials. We made spears, knives, bows, and arrows. With the weapons we made, we defended ourselves from the dangers in the new world, like monsters.

At least we have anti-radiation suits,

but we don't know how long will they protect us for. On his way to pick up some more steel from outside, Nathan heard a rustling noise. "Who's there?" He asked.

Yelled Nathan. Out of the rubble came out a dog-like entity. Nathan brought out his sword. The monster stuck out its tongue and made a panting sound. Nathan asked it, "Are you... friendly?"

The dog-thing did not make any attempts to answer back. Nathan put down his sword and started torward the dog-thing. Jumped into into Nathan's arms and started licking him.

"Aww, you're a cute little guy, aren't you?" said Nathan.

"I'm going to take you back to base. Okay?" the dog-thing barks.

On the way back though, the dog-thing goes into a barking fit. "What's wrong?" Nathan asks. It looks to the left. "Oh," Nathan says when he notices that there is a 10 foot monster.

The case shows that the expansion of forms of participation (i.e., from taking the role of designer to taking the role of designer and writer) could open the door for deepening the depth of participation and increasing the breadth of participation. While revising Brandon's texts as a writer, Nick changed from an audience to a contributor in writing. After that, he moved to a more central participation situation, being an originator while developing his own textual narratives. Brandon's (the writer in the team) interview responses further evidenced that Nick originated various story ideas, "first, we always talked about what will happen next, and then just started adding some implements, like he added the dog (blue animal in Figure 5) and then I added other ideas." The expansion to an originator also leads to a marked increase in the breadth of participation in writing. Nick developed many more textual narratives in the fall than in the summer.

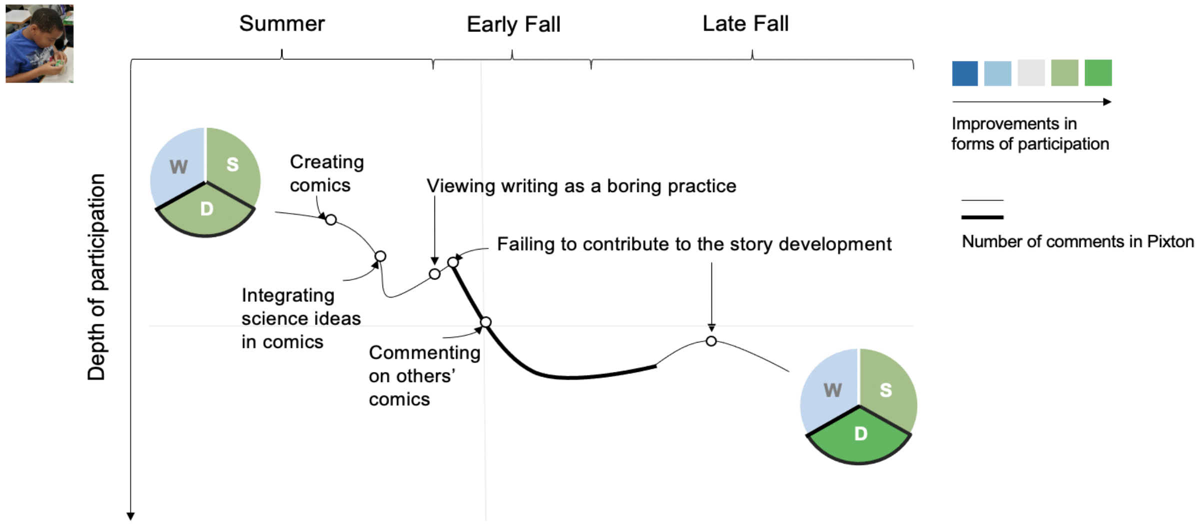
In summary, Nick's case demonstrates how role taking could be used as a tool to extend his comfort zone of design, helped him develop a strong interest in writing, and motivated him to take the role of writer. His diligence in design and helping team members ensured active participation in integrated disciplinary practices that connected different areas of disciplinary knowledge. The case also suggests that infusing himself in the story might motivate Nick to propose his own story ideas.

5.3. The case of Steve: A trajectory from creative designer to active commenter

Steve (sixth grade; African American male) took the role of designer in the team composed of one designer (himself), one scientist (Kaylee, seventh grade; African American female), and one writer (Pi, eighth grade; Indian American male). The team created a multimodal science fiction titled "Captain Atomicon." The team's multimodal composing processes involved the design of comics (Steve's major contribution), the development of textual narratives (Pi's major contribution), and the creation of a book cover (Pi's and Kaylee's major contribution).

Steve mostly worked alone during physical sessions but had active participation in online interactions. Most of the time when working on their team project, he worked by himself searching for online sources (e.g., images, videos, and music) and designing comics. In contrast to his working-alone style during small group meetings, he was very active in posting comments on others' comics in Pixton. Figure 6 summarizes Steve's participation in this program. In his case, forms of participation changed in the dimension of disciplinary practices while degrees of participation increased slightly in both breadth and depth.

Figure 6. Steve's participation trajectories

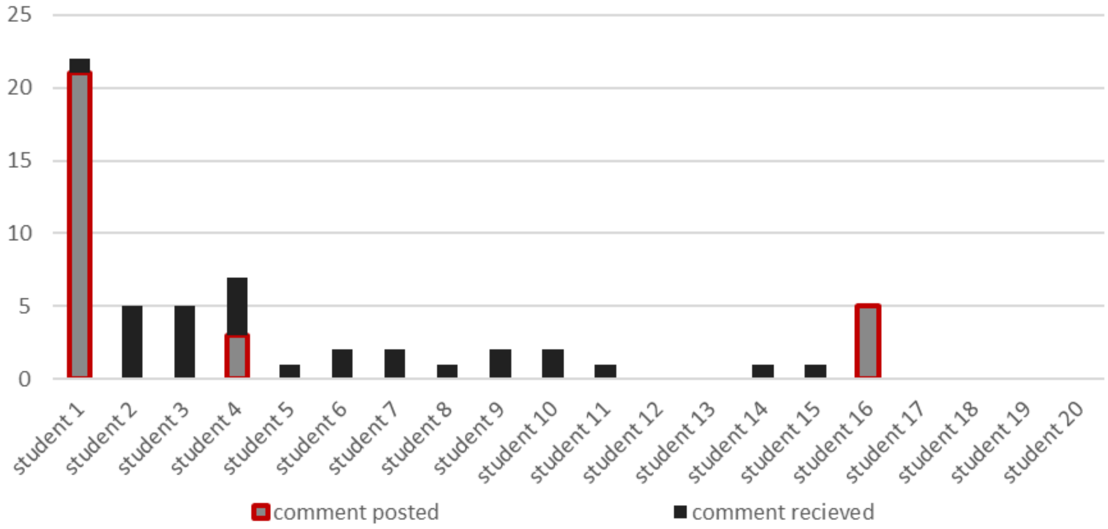


5.3.1. Forms and degrees of participation

With respect to forms of participation, Steve took on the role of designer, contributed to the team project as what the role entailed, and liked this role throughout, despite the fact that he objected to having specific roles. He demonstrated in his comics a better practicing designer over time (Figure 6). The improvement was evidenced in his more advanced comic design (e.g., better thematic congruence between comic panels), incorporating better writing in comics (e.g., speech bubbles with better writing), and integrating science ideas with multiple modes (e.g., using a combination of texts and pictures to illustrate science ideas). These different aspects of improvement were all grounded in his contribution as the designer of the team.

His breadth of participation was extended through the online space. Even though Steve's contribution to their team project was roughly evenly distributed throughout the sessions, over time, he posted more comments regularly to the whole class in Pixton. In the project, we did not prompt or require students to make comments so only a few (students 1, 4, and 16) made comments. Steve (Student 1 in Figure 7) posted the most comments.

Figure 7. Pixton activity from cohort B. Student 1 represents Steve



As we can see, although being silent and limited in physical interactions, Steve became more active in interacting with others in the community by posting online comments.

In terms of the depth of participation, Steve changed from more peripheral to more central participation through both contributions for the team project and the whole class. In the small group, he started with being an audience to try to understand Pi's fiction plot and a contributor to visualize the story. But gradually, he proposed new ideas in his comics. Specifically, he generated story plots that were sometimes complementary to but sometimes in conflict with Pi's textual narratives. While posting comments to the whole class, he contributed ideas instead of purely evaluating ideas from the aspect of design and science, but served more as an audience in writing. It indicates that he became a more central participant in design and science, but stayed as a more peripheral participant in writing.

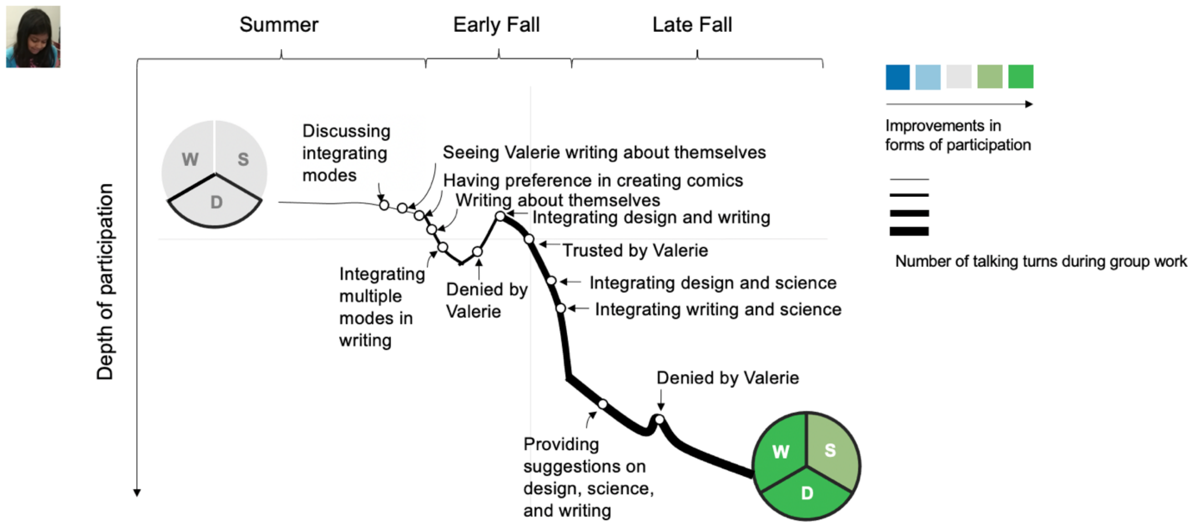
In summary, changes in the form of participation are reflected in the improvement in knowledge and practices associated with design while the degree of participation varied depending on the type of practices and the interactional spaces (i.e., face-to-face group interactions and online interactions in Pixton).

5.4. The case of Saanvi: A trajectory from conforming designer to proactive writer

Saanvi (Designer; fifth grade; Asian female) worked in a group with three sixth graders, including Valeria (Writer; Latina), Emilia (Scientist; Female; Other), and Mariana (Designer; Latina). The group created a multimodal science fiction titled "Tsunami Terror" in which Crystal, Sara, and Keke saved people from a tsunami that was triggered by Kai. The four main characters represented the four group members. While composing the fiction, Saanvi initially worked as a designer, but ended up taking all roles, especially designer and writer.

In the summer, Saanvi had remained on the periphery of the group but fulfilled her role of designer in spite of her silence and apparent isolation on the outer stage. She became strongly interested in taking the role of writer and reported the passion to take the role of writer on the last day of summer camp. In the second fall session, Saanvi had the chance to write the story while Valeria was absent. However, Valeria declined her changes to the story after coming back in the third session. Therefore, Saanvi went back to designing. While Valeria stuck with moving the storyline forward, Saanvi proposed to write narratives based on the comic that she created. Afterward, Valeria accepted Saanvi's writing. After building trust with Valeria, Saanvi had the opportunity to take the role of writer and put much effort into both writing and designing in the late fall. In addition, other group members trusted and valued Saanvi's suggestions while constructing artifacts towards the end of the project. Figure 8 summarizes Saanvi's participation in the STEM+L project. In this case, Saanvi had improvements in all the forms of participation and increases in both the breadth and depth of participation.

Figure 8. Saanvi's participation trajectories



5.4.1. Forms and degrees of participation

In terms of forms of participation, Saanvi's role changed from taking the role of designer to taking all three roles (Figure 8). Her change in forms of participation was mainly triggered by three factors: modal preferences, writing about themselves in the story, and integrating writing and science. Having preferences in designing specific modalities (e.g., comic), Saanvi became interested in taking the role of designer. In addition, she developed a strong interest in taking the role of writer after the summer camp. She wanted to write about herself, team members, and group interactions. This indicated that projecting themselves into the story through characters had the potential of fostering interest in writing. Although having a more negative attitude towards taking the role of scientist, she became interested in specific science topics in the story (e.g., tsunami). Her motivation for exploring specific science topics was to write the story in a more scientific way.

Her breadth of participation increased through composing with multiple modes across disciplines. Initially, Saanvi only created comics and, in her opinion, having multiple modes meant doing the same thing repeatedly. She gradually learned that modes could support each other in different ways. For example, she created voice narrations of science explanations to balance the fantasy aspect of textual narratives and science components. She described, "Fantasy is easier to do but with the science you have to have a little bit of equableness" (final interview). This case shows that multimodal composing across multiple disciplinary practices could extend the breadth of participation through composing with multiple modes and creating cross-disciplinary artifacts.

As made evident in this case, her depth of participation changed from more peripheral to more central participation after the trust between Saanvi and Valerie was established. Saanvi started with designing comics to visualize Mariana's ideas as a contributor. She also recorded voice narrations to explain the key science aspect of the story as a contributor. Over time, she could use modes that she created before as a way to originate ideas in another mode, story writing. For example, she expanded dialogues in comics into textual narratives when Valerie ran out of ideas. But her changes in writing were not accepted in the artifact until Valerie trusted and valued her writing. This case illustrates that being a contributor in one mode could lead to becoming an originator in another mode. Her more central participation was also evidenced in more active participation in group discussions. Being trusted as the one who contributed in developing story ideas, Saanvi provided ideas on all aspects of group artifacts in the late fall. Her participation was spread across various disciplinary practices when group members sought feedback from her.

In summary, Saanvi started as a quiet, passive participant but eventually ended up being a confident designer and writer who worked across multiple disciplinary domains, and an active team member others trusted and sought help. The case suggests that a student might contribute more (both in breadth and depth) in communities of practice when exposed to multiple disciplines and had flexibility in moving across disciplines to express oneself.

In addition, the extension of degrees of participation, mediated by composing with multiple modes, led to self-recognition as disciplinary persons. In this case, Saanvi clearly regarded herself as a designer, scientist, and writer who composed across disciplines (i.e., design, science, and writing) with multiple modes (e.g., comics, texts, and voice narrations). It indicates that multimodal composing might facilitate Saanvi in linking practices with disciplinary identities.

6. Discussion and implications

This study integrated the theoretical perspectives of disciplinary identity development and community of practice to examine students' participation from two dimensions: the form and the degree of participation. When comparing and contrasting these cases, commonalities and variations in changes in forms and degrees of participation were revealed. These commonalities and variations have implications for advancing our understanding of adolescents' participation trajectories in collaborative multimodal composition.

The four cases were similar in that students chose roles that fell within their practice comfort zones (Anderson & Gegg-Harrison, 2013). As a change in the form of participation, the four cases started with participating in disciplinary practices that they felt comfortable with (i.e., practice comfort zone) and used it as a bridge to move toward new disciplinary practices and take new roles. The cases reveal that students needed to see connections between disciplinary practices to move beyond practice comfort zone (Zimmerman & Land, 2014). Meanwhile, the extension of practice comfort zones set a critical foundation for the development of integrated disciplinary knowledge and practices as a change in the form of participation. Clearly, we can see that the four cases developed knowledge and skills in integrating different disciplinary knowledge and practices. This finding supports the current understanding of multiple modes driving disciplinary learning (e.g., Krajcik & Sutherland, 2010) and provides a new understanding of how students engaged in cross-disciplinary learning through composing with multiple modes.

While the four cases exhibited different breadth of participation across the program, all of them changed from more peripheral to more central participation - from consumers to producers. Scholars emphasized that learners should be encouraged to become producers of knowledge (Engle & Conant, 2002). This study adds to the research demonstrating the nuanced process of students changing from consuming ideas to producing ideas. All four cases ended up originating ideas, but the stimuli that triggered their changes and the timing when changes happened were different. Exploring the stimuli in other contexts is a promising future research direction to generate insights towards preparing students to be active learning agents in multimodal learning environments.

This paper also contributed in examining participation by connecting the form of participation, from the perspective of disciplinary identity and the degree of participation, from the perspective of community of practice. Firstly, composing with multiple modes helped grow (especially integrated) disciplinary knowledge. Improvements in disciplinary knowledge and practices provide fertile ground for the cultivation of disciplinary identities. This finding confirmed previous research indicating that multimodal composition could shape STEM identity (Pytash et al., 2017). Secondly, the interaction between self and the community has an impact on the construction of disciplinary identities. The recognition of disciplinary persons strengthens students' participation in sharing knowledge within and beyond small groups, and (even) beyond the community that students were in. Lastly, a close relation between self and artifacts, in the format of projection, motivates the extension of both breadth and (especially) depth of participation. It implies the crucial role of establishing the relation in moving students from knowledge consumers to producers. Further research is needed to probe the nuanced processes of building a close relation between self and multimodal artifacts and identify tools, materials, and activities to help students to connect themselves with artifacts.

In accordance with the literature (e.g., Hull & Katz, 2006; Smith et al., 2021), this study shows that students established a sense of ownership through composing with multiple modes. The sense of ownership can be strengthened with frequent within or across group sharing of multimodal artifacts. Furthermore, students demonstrated expertise in using composing tools. In a collaborative multimodal learning environment, the sense of ownership might create conflict. For example, in this study, Valerie initially denied Saanvi's contribution in writing. Conflict is a natural ingredient in any teamwork and avoiding conflict is not the solution (Jiang et al., 2021). Our study contributes to the literature by showing how students encountered and resolved conflicts (e.g., competing for using the same tool) while developing individual expertise. What needs to be explored further is instructional strategies that can turn these conflicts into opportunities to improve group performance. This includes transforming a competitive relationship into a collaborative one in which students could co-develop artifacts using the same mode and integrate different modes effectively. How to leverage the sense of ownership in multimodal composition to support teamwork would be a worthwhile research direction for the future.

The findings raise new questions about the conceptualization and operationalization of participation in STEM education. Students in this study had their own preferences in modes and disciplinary practices at the beginning of the project. Thus, their entry points in learning STEM practices were different. In addition, their preferred modes and practices could change dramatically over time. This phenomenon is in accordance with several studies in the literature (e.g., Nasir & Hand, 2008; Pinkard et al., 2017). It indicates that in operationalizing participation, we should consider students' previous experiences and interests, recognize that it's critical to capture dynamic changes, and acknowledge the great diversity of participation trajectories. Also, qualitative interpretation is a critical part of operationalization. One goal of operationalization is using the result of operationalization to design technologies for supporting teaching and learning activities (Boyce et al., 2014). One promising future direction is the study of showing the quantitative aspect of participation effectively for teachers to make qualitative interpretations.

The limitations of this study point to a number of important areas for future research. These findings are deeply situated in the STEM+L project where students created multimodal science fictions in small groups. Much more needs to be understood about participation trajectories with differing students, contexts, tools, and genres. Furthermore, the scope of this study was confined to how students participated in the project and did not capture aspects of their experiences that occurred outside of the project. Further research is needed that traces adolescents' participation across contexts and spaces.

Acknowledgement

This work was supported by the National Science Foundation under Grant 1713191. My deepest thanks to Drs. Ji Shen and Blaine E. Smith's support in data analysis.

Selection and participation of children

All consent processes and forms for this study were approved by the Solutions Institutional Review Board (IRB) (https://www.solutionsirb.com/) prior to the study's implementation. The parental consent forms were distributed and collected back before the study implementation. Students brought home the parental consent forms for signature. Parents provided consent to allow researchers to use student-generated data and conduct interviews.

References

Anderson, N., & Gegg-Harrison, T. (2013). Learning computer science in the" comfort zone of proximal development. In *Proceeding of the 44th ACM technical symposium on Computer science education* (pp. 495-500). https://doi.org/10.1145/2445196.2445344

Archer, L., DeWitt, J., Osborne, J., Dillon, J., Willis, B., & Wong, B. (2010). "Doing" science versus "being" a scientist: Examining 10/11-year-old schoolchildren's constructions of science through the lens of identity. *Science Education*, *94*(4), 617-639.

Boyce, C. J., Mishra, C., Halverson, K. L., & Thomas, A. K. (2014). Getting students outside: Using technology as a way to stimulate engagement. *Journal of Science Education and Technology*, *23*(6), 815-826.

Calabrese Barton, A., Kang, H., Tan, E., O'Neill, T. B., Bautista-Guerra, J., & Brecklin, C. (2013). Crafting a future in science: Tracing middle school girls' identity work over time and space. *American Educational Research Journal*, *50*(1), 37-75.

Carlone, H. B., & Johnson, A. (2007). Understanding the science experiences of successful women of color: Science identity as an analytic lens. *Journal of Research in Science Teaching*, *44*(8), 1187-1218.

Cobb, P., Confrey, J., DiSessa, A., Lehrer, R., & Schauble, L. (2003). Design experiments in educational research. *Educational Researcher*, *32*(1), 9-13.

Cohen, S. M., Hazari, Z., Mahadeo, J., Sonnert, G., & Sadler, P. M. (2021). Examining the effect of early STEM experiences as a form of STEM capital and identity capital on STEM identity: A Gender study. *Science Education*, *105*(6), 1126-1150.

Denault, A. S., & Poulin, F. (2009). Intensity and breadth of participation in organized activities during the adolescent years: Multiple associations with youth outcomes. *Journal of Youth and Adolescence*, *38*(9), 1199-1213.

Dou, R., & Cian, H. (2022). Constructing STEM identity: An Expanded structural model for STEM identity research. *Journal of Research in Science Teaching*, *59*(3), 458-490.

Engle, R. A., & Conant, F. R. (2002). Guiding principles for fostering productive disciplinary engagement: Explaining an emergent argument in a community of learners classroom. *Cognition and Instruction*, *20*(4), 399-483.

Erlandson, D. A., Harris, E. L., Skipper, B. L., & Allen, S. D. (1993). *Doing Naturalistic Inquiry: A Guide to Methods*. Sage.

Flyvbjerg, B. (2006). Five misunderstandings about case-study research. *Qualitative Inquiry, 12*(2), 219-245.

Franz-Odendaal, T. A., Blotnicky, K., French, F., & Joy, P. (2016). Experiences and perceptions of STEM subjects, careers, and engagement in STEM activities among middle school students in the maritime provinces. *Canadian Journal of Science, Mathematics and Technology Education*, *16*(2), 153-168.

Hall, T., & Coyne, P. (2005). *Narrative science writer phase II*. CAST.

Handley, K., Sturdy, A., Fincham, R., & Clark, T. (2006). Within and beyond communities of practice: Making sense of learning through participation, identity and practice. *Journal of Management Studies*, *43*(3), 641-653.

Hrastinski, S. (2008). What is online learner participation? A Literature review. *Computers & Education*, *51*(4), 1755-1765.

Hull, G. A., & Katz, M. (2006). Crafting an agentive self: Case studies of digital storytelling. *Research in the Teaching of English, 41*(1), 43-81.

Jiang, S., Smith, B. E., & Shen, J. (2021). Examining how different modes mediate adolescents' interactions during their collaborative multimodal composing processes. *Interactive Learning Environments, 29*(5), 807-820.

Jordan, B., & Henderson, A. (1995). Interaction analysis: Foundations and practice. *The Journal of the Learning Sciences*, *4*(1), 39-103.

Krajcik, J. S., & Sutherland, L. M. (2010). Supporting students in developing literacy in science. *Science*, *328*(5977), 456-459.

Kress, G. (2003). *Literacy in the new media age*. Routledge.

Kress, G., & Van Leeuwen, T. (2001). *Multimodal discourse: The Modes and media for contemporary communication*. Edward Arnold.

Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation*. Cambridge University Press.

Malinen, S. (2015). Understanding user participation in online communities: A Systematic literature review of empirical studies. *Computers in Human Behavior*, *46*, 228-238.

Miles, M., & Huberman, M. (1994). *Qualitative data analysis*. Sage.

Nasir, N. I. S., & Hand, V. (2008). From the court to the classroom: Opportunities for engagement, learning, and identity in basketball and classroom mathematics. *The Journal of the Learning Sciences*, *17*(2), 143-179.

National Research Council (2017). *Communicating science effectively: A Research agenda*. National Academies Press.

NGSS Lead States. (2013). *Next Generation Science Standards: For states, by states*. The National Academies Press.

Patton, M. Q. (1990). *Qualitative evaluation and research methods*. Sage Publications.

Pinkard, N., Erete, S., Martin, C. K., & McKinney de Royston, M. (2017). Digital youth divas: Exploring narrative-driven curriculum to spark middle school girls' interest in computational activities. *Journal of the Learning Sciences*, *26*(3), 477-516.

Preece, J. (2001). Sociability and usability in online communities: Determining and measuring success. *Behaviour & Information Technology*, *20*(5), 347-356.

Pytash, K. E., Kist, W., & Testa, E. (2017). Remixing my life: The Multimodal literacy memoir assignment and STEM. *Journal of Adolescent & Adult Literacy*, *61*(2), 163-172.

Smith, B. E. (2014). Beyond words: The Landscape of research on adolescents and multimodal composition. In K. E. Pytash & R. E. Ferdig (Eds.), *Exploring Multimodal Composition and Digital Writing* (pp. 1-19). IGI Global.

Smith, B. E., Beach, R., & Shen, J. (2021). Fostering student activism about the climate crisis through digital multimodal narratives. *Journal of Sustainability Education*, *25*. http://www.susted.com/wordpress/content/fostering-student-activism-about-the-climate-crisis-through-digital-multimodal-narratives_2021_08/

Smith, B. E., & Dalton, B. (2016). Seeing it from a different light. *Journal of Adolescent & Adult Literacy*, *59*(6), 719-729.

Smith, B. E., & Shen, J. (2017). Scaffolding digital literacies for disciplinary learning: Adolescents collaboratively composing multimodal science fictions. *Journal of Adolescent & Adult Literacy*, *61*(1), 85-90.

Spradley, J. P. (1980). *Participant observation.* Holt, Reinhart & Winston.

Strauss, A., & Corbin, J. (1998). *Basics of qualitative research techniques*. Sage Publications.

Van Horne, K., & Bell, P. (2017). Youth disciplinary identification during participation in contemporary project-based science investigations in school. *Journal of the Learning Sciences, 26*(3), 437-476.

Wenger, E. (1998). *Communities of practice: Learning, meaning, and identity*. Cambridge University Press.

Zheng, B., & Warschauer, M. (2015). Participation, interaction, and academic achievement in an online discussion environment. *Computers & Education, 84*, 78-89.

Zimmerman, H. T., & Land, S. M. (2014). Facilitating place-based learning in outdoor informal environments with mobile computers. *TechTrends, 58*(1), 77-83.

Jiang, M. Y. C., Jong, M. S. Y., Lau, W. W. F., Chai, C. S., & Wu, N. (2023). Effects of Automatic Speech Recognition Technology on EFL Learners' Willingness to Communicate and Interactional Features. *Educational Technology & Society*, *26*(3), 37-52. https://doi.org/10.30191/ETS.202307_26(3).0004

Effects of Automatic Speech Recognition Technology on EFL Learners' Willingness to Communicate and Interactional Features

Michael Yi-Chao Jiang[1], Morris Siu-Yung Jong[2,3], Wilfred Wing-Fat Lau[2], Ching-Sing Chai[2] and Na Wu[4*]

[1]School of Foreign Languages, Shenzhen Technology University, China // [2]Department of Curriculum and Instruction, The Chinese University of Hong Kong, Hong Kong SAR // [3]Centre for Learning Sciences and Technologies, The Chinese University of Hong Kong, Hong Kong SAR // [4]College of International Business, Shenyang Normal University, China // mjiang@sztu.edu.cn // mjong@cuhk.edu.hk // wwflau@cuhk.edu.hk // cschai@cuhk.edu.hk // wunacib@synu.edu.cn

[*]Corresponding author

ABSTRACT: This study examined the effects of using automatic speech recognition (ASR) technology on Chinese students' willingness to communicate (WTC) in oral English and the development trajectories of their interactional features in a flipped EFL context. One hundred sixty undergraduates from a Chinese university participated in the 14-week quasi-experiment. Both groups were taught in a flipped fashion. The treatment group was required to use the ASR technology for oral practice in their pre-class self-learning, while the control group conducted their self-learning without the ASR technology. The results found that the ASR-based oral practice led to a significant between-group difference in students' WTC with teacher and class and WTC with non-Chinese, showing that the ASR technology may contribute to improving the Chinese students' WTC in oral English. Conversely, except for the between-group effect on negotiation for meaning, there was no significant difference between the two groups on the other measures of interactional features. Moreover, none of the interactional features of the students in the treatment group changed significantly over time, indicating a limited role of the ASR technology on Chinese students' interactional features. Discussions were conducted regarding the contradictory effects of the ASR technology on WTC and peer interaction.

Keywords: Automatic speech recognition, English as a foreign language, Interactional features, Willingness to communicate

1. Introduction

Automatic speech recognition (ASR) is known as a computer-based process of decoding and transcribing oral language usually into text form (Kim, 2006). It is a specialized application of artificial intelligence in natural language processing and has been broadly incorporated into various scenarios in modern daily life (Evers & Chen, 2020). In recent years, free speech-to-text ASR-based technology (i.e., dictation ASR) has gained growing attention in the domain of foreign language (FL) education. Owing to the real-time feedback on FL learners' oral performance without time and space restrictions (Wang & Young, 2014), mobile-based dictation ASR technology can be integrated into out-of-class self-learning to enhance students' preparedness for in-class activities (Jiang et al., 2022a). As such, flipped classrooms that realize the switch of in-class lectures and out-of-class assignments seem to be "a good match" for dictation ASR technology in FL learning (e.g., Jiang et al., 2021a). In view of Bloom's taxonomy (Anderson & Krathwohl, 2001), the in-class activities in a flipped classroom are mostly for developing higher-order thinking skills in terms of applying, analyzing, evaluating, and creating (Jong, 2019). Contrastingly, the out-of-class learning is remembering- and understanding-oriented (Jong, 2017; Jong et al., 2019). Bloom's taxonomy contextualized into flipped FL classrooms, the in-class activities are generally composed of output-oriented tasks that can elicit language use in learners collaboratively, while the out-of-class self-learning abounds with input-oriented content knowledge learning such as vocabulary and reading comprehension.

Typically, under the influence of Chinese culture which advocates learners' conformity (Lee & Yin, 2011), Chinese students are often portrayed as obedient learners who are reluctant to express opinions. Particularly, in an English as a foreign language (EFL) classroom, probably due to their limited English, Chinese students do not seem to have strong willingness to communicate (WTC) in participating in in-class activities (Lou & Noels, 2021). Their peer interactions are typically limited to short phrases and even single words (Clark & Gieve, 2006). Tang et al. (2020) argued that the Chinese learners' reticence was a learnt behavior driven by intentions consisting of behavioral, normative and control beliefs towards classroom participation. While not all the students possess homogeneous learning styles (Tran, 2013), some studies have claimed that Chinese students are not accustomed to participating actively in the classroom (e.g., Tang et al., 2020). They argued the traditional

Chinese teaching and learning atmosphere did not create situations where raising questions and challenging peers was regarded as necessary (Durkin, 2011).

Owing to its flexibility, dictation ASR technology can facilitate free and unguided speech practice and allows users to dictate anything in the target language that interests them (Evers & Chen, 2020). In other words, it affords learner autonomy which is generally conducive to learning (Ryan & Deci, 2020). In addition, dictation ASR allows flexibility in the models of speech and students may not get discouraged if they speak with a different English accent or variety (Hincks, 2015). Furthermore, dictation ASR can be better integrated into the design of pedagogical activities, as "teachers can assign practice with sound or topics that they have been working on in class instead of sending students off to follow a program that may not correspond clearly with the class." (McCrocklin, 2019, p. 103) Those benefits of dictation ASR are thought to better facilitate students' pre-class self-study in an EFL class, thus resulting in a higher level of learner preparedness (Jiang et al., 2022a). Consequently, it is hypothesized that the ASR-based oral practice is likely to boost students' willingness to interact with their peers in class. On the other hand, because of the ample opportunities of speaking practice assisted by dictation ASR, the practice effect may have a significant impact in that students' speaking anxiety may be reduced and self-efficacy in speaking may be increased. They may, therefore, be more willing to conduct in-class peer interaction in the target language.

Although researchers have revealed that ASR-based technology may enhance users' EFL learning in terms of affective and behavioral variables, (e.g., Ahn & Lee, 2016; Jiang et al., 2022b), few studies examined the effects of ASR-based technology on learners' EFL learning from the perspective of peer interaction, and little is known about whether the ASR-based technology may contribute to the development of students' WTC and peer interaction. In a flipped FL classroom, the in-class activities are designed to be highly interactive featuring higher order language skills (e.g., analyzing- and evaluating-oriented tasks). However, Chinese learners' reticent non-participating behaviors in class might lead to insufficient in-class interactions (Jiang et al., 2022a). Hence, there is a need to raise the effectiveness of the ASR-based technology for enhancing peer interaction in the Chinese EFL context. Therefore, the current study was conducted to examine the effects of the ASR-based technology on Chinese students' WTC and peer interaction in a flipped EFL classroom.

2. Related works

2.1. ASR-based technology in flipped language classrooms

Intelligent computer assisted language learning (iCALL) alongside online resources can promote learner autonomy "by enabling experimentation through self-access work outside of class" (McCrocklin, 2016, p. 27). Compared with practicing with native speakers or language teachers, iCALL practice is not subject to time and space constraints (Jiang et al., 2021a), thus providing ample opportunities for developing learner autonomy. Pedagogically, such advantages of ASR coincide with the tenets of flipped learning for FL classrooms. The pre-class self-study in flipped learning requires adequate learner autonomy, indicating that the student should play a role of autonomous learners taking full charge of their learning. In other words, the students need to take responsibility for all the decisions in relation to their pre-class self-learning. The ASR technology transcribes students' speech into text, providing immediate and more importantly, visualized feedback on their speech (Levis, 2007), and students can see immediately what they say in the target language. In flipped language classrooms, therefore, the ASR-based technology is assumed to help students monitor their flow of speech and support to self-assess their output.

Learner preparedness is identified as a crucial factor for the success of the flipped classroom approach (Sun & Xie, 2020), and with the ASR-based technology, students in a flipped FL classroom are expected to be prepared with a better mastery of the contents for in-class higher order activities. Consequently, those with higher level of learner preparedness tend to be more willing to communicate with their peers. According to Tai and Chen (2020), interaction with the ASR-based application enhanced EFL learners' confidence in communication and reduced their speaking anxiety, resulting in a higher level of WTC. Furthermore, they also claimed such positive effects of ASR-based technology on WTC were attributable to increased interaction and engagement, and the less threatening environment the ASR-based application established. Overall, in flipped FL classrooms, the ASR-based technology provides an avenue for developing learner autonomy and enhances students' pre-class preparedness, which in turn may boost their WTC and ready them to engage in more peer interaction while performing higher-order tasks in class.

The benefits of ASR-based technology consist in the significant amount of practice, consistency, the unbiased nature of feedback and diverse forms of visual representations (Levis, 2007). Especially, the real-time assessment of learners' utterances is beneficial for acquiring listening and speaking skills (Wang & Young, 2014). Recently, a growing number of empirical studies have obtained positive evidence of ASR-enhanced FL learning (e.g., Ahn & Lee, 2016; Jiang et al., 2021a). It was concluded that ASR-based technology could provide language learners with (1) more opportunities to produce and extensive interaction in the target language, (2) immediate feedback in various direct and indirect forms, and (3) more control over their learning with increased confidence. On the learners' part, ASR-based technology can create a less threatening self-paced environment for them to learn to speak the target language (McCrocklin, 2016).

However, a closer look at those studies reveals that most of them relied on participants' affective and behavioral data to evaluate the effectiveness of ASR-based technology, and few studies investigated the potential changes from the interactional features of peer interaction. For example, Ahn and Lee (2016) investigated junior students' attitudes towards the use of an ASR-based application using close- and open-ended questionnaires. Evers and Chen (2020) surveyed adult EFL learners' learning styles through questionnaires and used human rating method to evaluate the effects of an ASR system on students' pronunciation performance. Although they are widely used in FL educational research, affective and behavioral data are mostly self-reported and subjective, which tend to be biased to some extent (Wilson & Zietz, 2004). In that case, the effectiveness of ASR-based technology on students' language performance may not be sufficiently evaluated. Moreover, the social dimension is also crucial in understanding the impact of ASR-based technology on students' FL learning. Indicators such as interactional features of the participants' peer interaction can provide a more comprehensive perspective of explaining students' language learning and may contribute to diversifying FL pedagogy.

2.2. WTC in EFL classrooms

In the domain of language learning, some researchers regard WTC to be a personality trait, while by many others WTC is taken as a context-dependent domain-specific variable (Tavakoli & Davoudi, 2017). Despite the complex nature of WTC which manifests itself in its diverse conceptualizations (Pawlak & Mystkowska-Wiertelak, 2015), it is one of the crucial determinants of language classroom communication (Tavakoli & Davoudi, 2017). Generally, WTC is conceptualized as a readiness to speak in an FL or second language (L2) in a particular situation with a specific person (MacIntyre, 2007) and thus perceived as part of the broader concept of FL learning motivation. As the "final psychological step to the initiation of L2 communication" (MacIntyre & Doucette, 2010, p. 161), WTC can indicate how well the students are engaged in the collaborative tasks in EFL classrooms on the social dimension.

A plethora of primary studies have obtained empirical evidence that WTC is highly associated with FL learners' communication in the target language (e.g., Zhang et al., 2020). Factors such as individual's language and communicative competence, affective and environmental factors were reported to have significant impact on language learners' WTC in various pedagogical contexts. For example, Sato (2020) revealed that individuals' target language proficiency might moderate the impact of various factors on their WTC. Low-intermediate speakers' WTC was more likely to be affected by their interest in the topic and self-confidence, whereas advanced speakers' WTC was influenced by the opportunity to talk about themselves and their opinions. Wang et al. (2019) found that learners' perception of group interaction and interaction with the language teacher had a significant effect on students' WTC and their classroom communication in the target language. Zare et al. (2020) investigated the interplay of oral corrective feedback (explicit correction, recasts, and prompts) and L2 WTC across different L2 proficiency levels. They found that learners preferred prompts most and elicitation-oriented feedback were the most contributory to L2 WTC.

While numerous empirical studies have been conducted to investigate the interplay of WTC and other relevant factors in traditional EFL classrooms, little is known in literature about how ASR-based technology could influence learners' WTC in EFL classrooms. Given the Chinese students' typical reticence in class, it is worth investigating whether the ASR-based technology could enhance their WTC for in-class peer interaction.

2.3. Peer interaction in flipped EFL learning

In a student-centered language classroom, peer interaction or learner-learner interaction is the most significant interaction that leads to language learning and development. Therefore, how students interact to co-construct meaning and knowledge of language form is a crucial issue in FL learning research (Loewen & Sato, 2018). Ever since the formulation of the interactionist hypothesis (Long, 1981), which claims that interaction, particularly when it involves negotiation for meaning and feedback, facilitates L2 acquisition, interactionist approach has long been considered a theoretical underpinning for task-based language instruction. Numerous empirical studies based on the interactionist approach have provided “the main psycholinguistic underpinnings for task-based language teaching” (Long, 2015, p. 61). To date, studies on task-based peer interaction in language learning have realized that interactional features provide learning opportunities that are theoretically posited to be beneficial to EFL learning. As a review study found (Mackey & Goo, 2007), the occurrence of a variety of interactional features in task-based peer interaction can facilitate second language acquisition. In the past decade, a growing number of empirical EFL studies have investigated a range of interactional features (e.g., clarification requests, recasts and uptakes) and their effects on students’ language learning.

Following Philp et al. (2014), peer interaction provides language learners “a context for experimenting with the language” (p. 17) and tends to be less threatening than teacher-led interaction, because students do not have to worry about making errors while talking to their peers. As with the other types of interaction in a language classroom, peer interaction is comprised of input, negotiation, output and noticing (Loewen & Sato, 2018). Correspondingly, a range of interactional features are established by interactionist researchers to operationalize those four constructs. Negotiation for meaning (NfM) is at the core and acts as a response to a communication breakdown, including clarification requests, confirmation checks and comprehension checks (Pica, 1994). When they cannot get themselves across, language learners may signal that a communication breakdown has occurred. However, NfM does not always occur frequently in classroom contexts. L2 or FL learners also have negotiation of form as a result of a desire for linguistic accuracy rather than as a result of communication breakdown (Gass et al., 2011). They may talk about the language they are producing or question their language use, which was defined by Swain and Lapkin (1998) as language-related episodes (LREs). LREs are identified in a myriad of primary studies as learning opportunities during interaction. One particular type of input-providing LRE is recast, which is a “correct restatement of a learner’s incorrectly formed utterance” (Nicholas et al., 2001, p. 720). A recast is didactic in nature and provides the correct linguistic form for the learner. It is often provided by the interlocutor who is more proficient in the target language as a type of corrective feedback. Extant studies have demonstrated that learners can both notice and benefit from recasts despite individual and contextual variability (e.g., Rassaei, 2022), indicating recast plays an indispensable role in the interactionist approach.

In EFL classrooms, collaborative tasks are typically designed to elicit peer interaction in the target language, wherein learners may notice the necessary language forms and focus on meaning. Compared with the activities in a lecture-based EFL classroom, the higher-order tasks in flipped EFL classrooms are usually more complex and students need to allot more attentional resources to perform those tasks. For example, those higher-order tasks may involve more elements and verb tenses or more perspective taking. Therefore, according to the Cognition Hypothesis (Robinson & Gilabert, 2007), those tasks would promote more interactional features, which in turn lead to language development. Consequently, more peer interaction could be witnessed when students are performing more complex and higher-order tasks.

However, in the eyes of most Chinese learners, only their teachers comment on their academic performance and provide them with corrective feedback. Such conventionally deep-rooted perceptions are closely associated with the key merit of rituals/etiquette (*Li*) in Chinese culture, which refers to the ethic of propriety and the prescription of social relationship structures (Yum, 2007). If they replaced their teacher’s role and provided feedback directly to their peers, the students would be considered as a “show-off” by their peers. Besides, acting as a teacher may result in an unnecessary deviation in power relations with their peers. Therefore, Chinese students psychologically resist playing a teacher’s role in front of their peers (Xu & Kou, 2018) and may only participate in a discussion when they have something “safe” to say (Wu, 2015). On the other hand, the long-held test-orientated learning may “push” the Chinese EFL learners to make every effort only to score (Gao, 2008), losing interest in mastering English as a means of communication. Most of them were so accustomed to the exam-oriented and didactic English teaching that they might find themselves speechless when they were assigned free-talking tasks in EFL classrooms.

This study aimed to examine the effects of the ASR-based technology on Chinese students’ WTC and peer interaction in a flipped EFL classroom. Specifically, it was conducted to (1) investigate whether the integration of the ASR-based practice into pre-class self-learning could contribute to the development of Chinese EFL students’ WTC and their interactional features, and (2) reveal the development trajectories of Chinese students’ interactional features over time in a flipped EFL classroom. Accordingly, three research questions were formulated:

RQ 1. Does the ASR-based technology lead to any difference in Chinese EFL learners’ WTC in a flipped classroom?

RQ 2. Does the ASR-based technology lead to any difference in Chinese EFL learners' interactional features in a flipped classroom?

RQ 3. How does the ASR-based technology influence the development trajectories of Chinese EFL learners' interactional features in a flipped classroom?

3. Methods

3.1. Participants and research background

Four classes of 160 freshmen from a public university in China consented to participate in this study. They majored in a range of disciplines, including mathematics, physics, education, psychology, Chinese language and arts, chemistry, and computer science. According to a pre-course survey, the participants had learnt English for approximately 10.9 years on average; 69.4% reported little or no training in oral English, and 92.5% had little or no experience of flipped learning. They all registered for the same EFL course, but according to the pre-course placement test, most were disfluent in oral English. The four classes were randomly assigned into a control group (CG) of two classes and a treatment group (TG) of the other two classes.

3.2. Instructional design

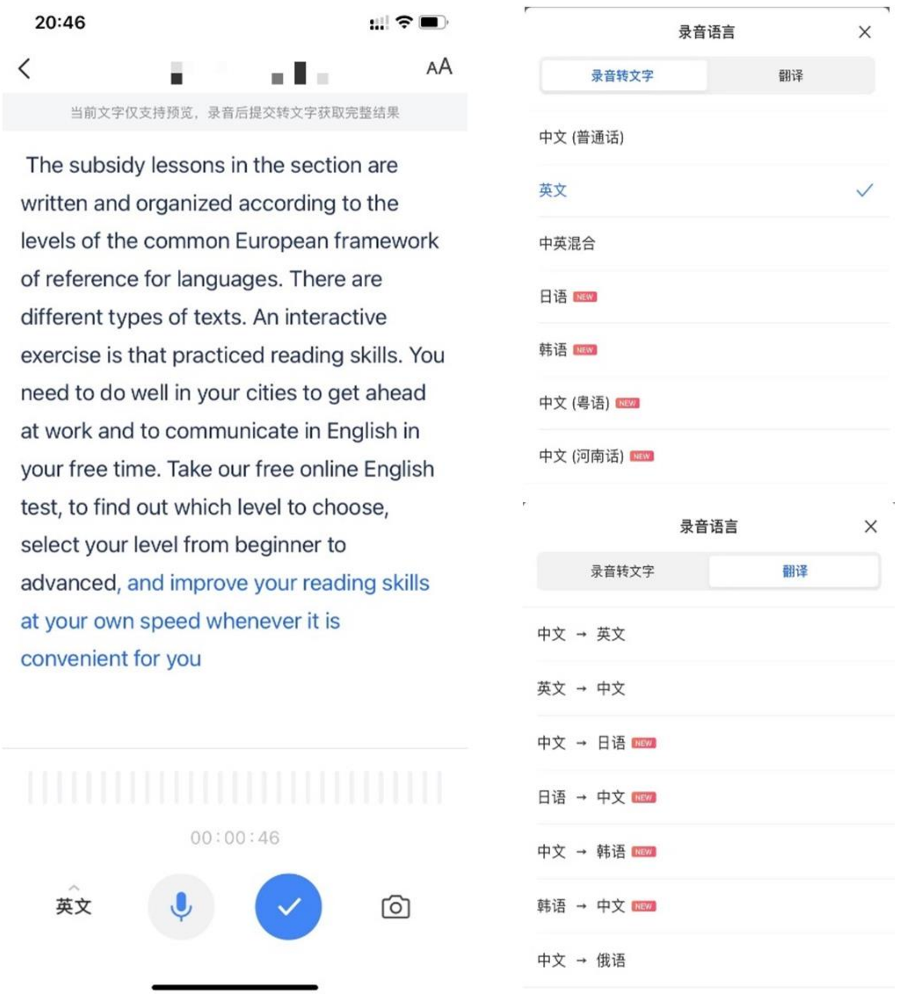
The course lasted for 14 weeks in the 2019 fall semester. It was a compulsory EFL course for Year 1 and Year 2 students which aimed to develop learners' English proficiency, especially to foster the students' oral language skills. Before the course started, a placement test was administered to measure students' overall English proficiency. The test was made up of a computer-based section (composed of reading comprehension, listening comprehension and essay writing) and an in-person individual interview. The computer-based section lasted for 100 minutes, and the oral interview was approximately five minutes per candidate.

Week 1 was the course orientation, and Weeks 2–13 covered the instruction of eight learning units. Each week, the students in both groups had one 90-minute face-to-face session. In Week 14, both groups took an end-of-semester test as a summative assessment of the course. Within each class, the students were randomly assigned into subgroups of three or four for in-class activities. Both groups were taught by the same teacher with a flipped classroom approach. An e-learning platform called Unipus (http://learn.unipus.cn/, Unipus https://u.unipus.cn/index.html/, is an online learning platform developed and owned by Foreign Language Teaching and Research Press in China, which provides hybrid teaching solutions for foreign language teaching in colleges and universities. This platform provides a one-for-all package of learning, practicing, testing, and evaluating that feature interactive experience.) was utilized to implement the flipped instruction. One week before each face-to-face session, the flipped materials were delivered to the students via Unipus for pre-class self-learning. In terms of Bloom's taxonomy (Anderson & Krathwohl, 2001), the flipped contents were remembering- and understanding-oriented and served as lead-ins for in-class higher order activities. Those materials were developed and organized in the form of a massive open online course by the textbook publisher and could be accessed conveniently by the students. The students were required to master the flipped content using Unipus and raise questions through WeChat (a social media/instant messaging platform) when needed. The textbook featured guided dialogues to enable participatory communication and interaction. The course teacher determined the learning materials from the textbook for both pre-class and in-class tasks. Generally, all the reading-focused sections alongside the vocabulary- and grammar-associated sections were "flipped" out of class for students' self-learning because they could serve as input to prepare the students for the in-class higher-order activities that concern the unit themes.

Particularly for the pre-class self-study, both groups were required to conduct a free-response oral practice to check the students' mastery level of the flipped content. A unit theme-focused free-response question was given to the students, and they needed to base their responses on their pre-class self-study. The students in both groups were encouraged to practice performing that oral task repeatedly. Then they needed to record and upload their responses to Unipus and their speech had to last for at least two minutes. The CG students performed this task with no additional requirements, while the TG students were required to utilize an ASR-based application called iFlyRec (https://www.iflyrec.com) to perform this task (Figure 1). This application is developed by iFlyTek, a renown artificial intelligence company, and it is free to download and can run on iOS and Android.

Contrary to computer assisted pronunciation training (CAPT) systems that are specifically designed for FL learners, iFlyRec is a mobile-based dictation ASR application originally developed for native speakers. Since they are not designed for pedagogical purposes, mobile-based dictation ASR applications do not provide analysis of speech or artificial intelligence-based interaction (e.g., Google Assistant) with users. However, because of its flexibility, dictation ASR applications would be more effective in FL teaching combined with scaffolded activities (Evers & Chen, 2020). More specifically, dictation ASR applications allow users to try practicing any word, phrase, or topic that interests them (Hincks, 2015), while CAPT programs are not able to facilitate free, unguided speech practice (McCrocklin, 2019). Besides, the flexibility of dictation ASR applications is also reflected in the models of speech (e.g., Received Pronunciation or American Standard), meaning ASR-based oral practice allows FL learners to speak with different accents and varieties. Apart from the above-mentioned advantages that all dictation ASR-based applications have, iFlyRec has some exceptional features that distinguish it from other dictation ASR-based applications. It realizes real-time conversion from speech to text across multiple languages and even some Chinese dialects. Additionally, it can also be used for interlingual translation based on ASR technology in several languages such as Chinese, English, and Russian.

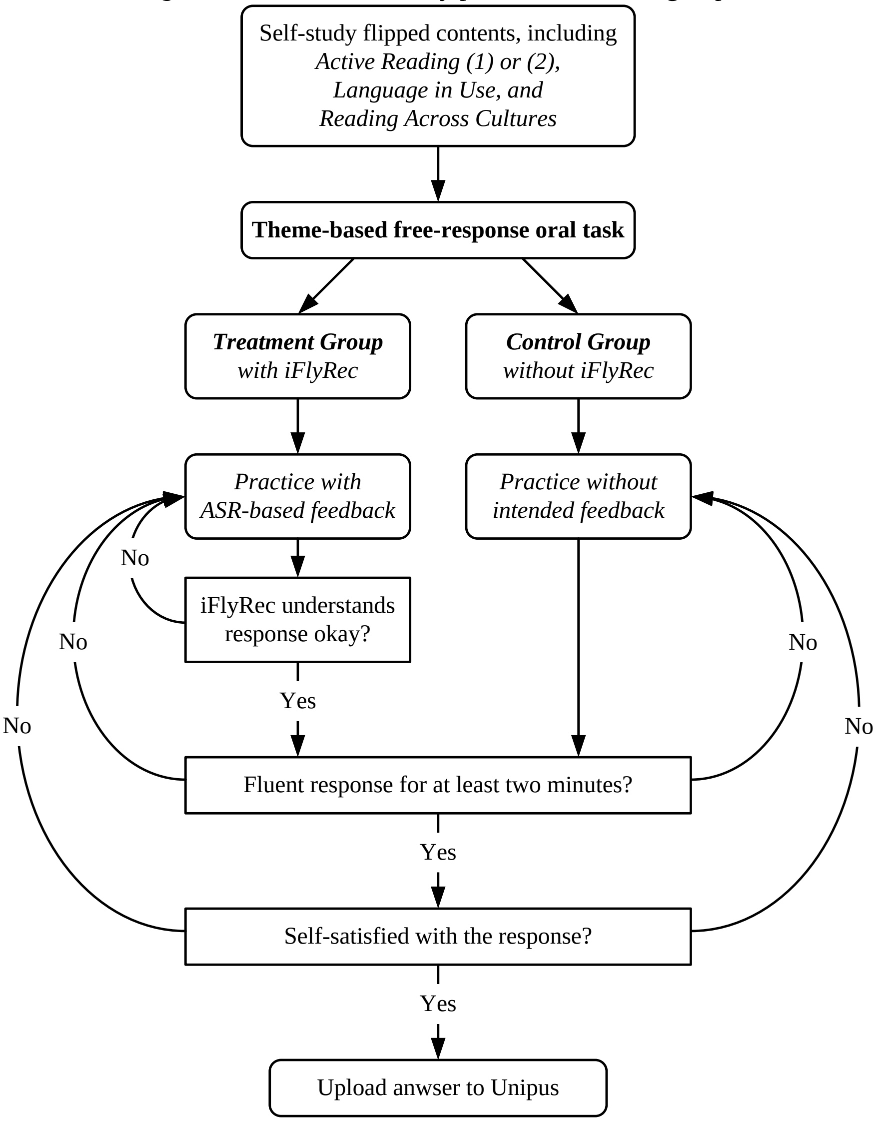
Figure 1. Screen capture of iFlyRec



When performing the ASR-assisted oral tasks, the TG students could refer to the transcribed texts as immediate feedback to easily identify morphosyntactic or pronunciation-related errors in their speech. Thus, the TG students could improve their task performance by correcting themselves until they could make themselves understood by the application fluently. Conversely, the CG students did not have any particular visualized feedback to support their task-performing. Since they were given one week's time to perform the oral task, both groups were recommended to keep on practicing until they were satisfied with their oral answers. (see Figure 2 for the procedures of pre-class self-study of both groups).

For each unit of in-class learning, the students in both groups undertook one shared Unit Task (UT), which was designed to elicit their language use in the classroom setting (see Figure 3 for a sample UT). They were allowed five minutes to communicate with their subgroup members and perform the task collaboratively. The students were required to record their performance using their mobile devices and upload their recordings to Unipus. Among the eight units, the recordings of Units 2, 4, 6 and 8 were chosen for data analysis, but the course teacher and the students were informed that recordings of all the eight units would be analyzed for the sake of their performance consistency.

Figure 2. Pre-class self-study procedures of both groups



3.3. Instrument and measures

3.3.1. Measuring WTC

To survey participants' WTC in oral English, a WTC questionnaire developed by Tavakoli and Davoudi (2017) was adopted in this study because it was specifically created for the Asian EFL learning context (Appendix A). To ensure respondents' full understanding of the items, the original questionnaire was translated from English into Chinese. Then, backward translation was conducted to guarantee the Chinese version was equivalent in meaning to the original questionnaire. A panel of three professors in the domain of FL education were consulted, and a pilot study (n_p = 16) was conducted among another homogeneous group of the participants. Modifications were made following the feedback from the panel and the pilot students before the questionnaire was finalized.

The questionnaire consists of 27 items, assessing WTC in oral English in three different contexts: WTC with teacher and class (13 items), WTC with friends (7 items) and WTC with non-Chinese (7 items). Reliability analysis of the data showed that the Cronbach α of the pre-intervention survey was 0.845, 0.789 and 0.744 on the three sub-questionnaires, and that of the post-intervention survey was 0.874, 0.795 and 0.816, respectively, indicating good internal reliability of both measurements.

Figure 3. A sample UT for data collection
Unit 4 *Family Affairs*

In-class Unit Task 1:

- ❑ **Make a list of the people in your family, and draw lines between them to show the relationships. Use photos if you have them.**
- ❑ **Now write a brief description of each person. Make sure you include some details about:**
  - their name
  - their age
  - their nickname (if they have one)
  - their character
  - how well you get on with them
  - any favourite or typical stories about them

↓

In-class Unit Task 2 (for data collection):

- ❑ **Invent a "false" member of your family (someone who doesn't exist). Write a brief description of them, and invent any details. Try to make them as real as possible.**
- ❑ **Work in pairs.**

  Student A: Talk about your family to Student B, including the "false" member.

  Student B: Try and decide which is the "false" member of Student A's family.

3.3.2. Measuring interactional features

A between- and within-subjects design was adopted to analyze the effects of the ASR-based technology on interactional features and the development trajectories. The independent variables were group (i.e., between-subjects factor) and time (i.e., within-subjects factor). The dependent variables were the participants' interactional features coded from their in-class task-performance at four timepoints (Units 2, 4, 6, and 8). Repeated measures ANCOVA was conducted to analyze the data. Following Gass et al. (2011), the current study operationalized interactional features as a set of measures, a practice that has been widely applied in interactionist studies. The measures include (a) NfM, which is further operationalized as clarification requests, confirmation checks and comprehension checks; (b) LREs, and (c) recasts (Appendix B).

The current study quantified each measure of interactional features in terms of their relative frequency, which was computed by dividing the frequency of a given measure over a production unit (Norris & Ortega, 2009). In this study, the analysis of speech unit (AS-unit) was chosen as the production unit of the participants' utterances. An AS-unit is "a single speaker's utterance consisting of an independent clause, or sub-clausal unit, together with any subordinate clause(s)" (Foster et al., 2000, p. 365). It was revised based on the extant production units available to particularly deal with the fragmentary nature of oral data and provide a solution to the fuzziness and complexity of the spoken language (Jiang et al., 2021a).

3.4. Data collection and analysis

The WTC questionnaire was administered right before and after the intervention. A total of 155 responses (out of 160) were collected because five participants were absent for either the pre- or the post-intervention survey. To minimize possible data contamination caused by "careless" respondents, following Jiang's et al. (2021b) method, we added three "filtering items" to the questionnaire which served as an indicator of the participants' consistency of their responses. Consequently, 24 participants were removed from the sample, leaving 131 participants for the analysis of the WTC data (68 CG students and 63 TG students). One-way ANCOVA was conducted to examine the effect of using ASR-based technology on participants' WTC in oral English. The participants' pre-intervention English proficiency and WTC in oral English were controlled for in the data analysis as covariates.

Because of classroom noise, dropout and absence issues, a sum of 32 participants were removed from the recordings sample, leaving 128 participants for the analysis of interactional features (60 CG students and 68 TG students). The recordings were transcribed verbatim and then coded with ELAN (https://archive.mpi.nl/tla/elan), a professional software for annotating audio and video data in linguistic studies. Two authors were involved in the coding process, and the inter-coder reliability estimated by Krippendorff's α (Hayes & Krippendorff, 2007) was calculated to be 0.818 (> 0.8), indicating that the co-coding was deemed consistent between the two coders. Any disagreement was resolved through discussion until an agreement was reached. Annotated examples of NfM, LRE and recast are provided in Appendix B.

4. Results

4.1. Change in WTC

Descriptive statistics showed that compared with their CG counterparts, the TG students had higher post-intervention WTC scores in all the three contexts (Table 1). One-way ANCOVA was then conducted with the participants' pre-intervention overall English proficiency and WTC scores as covariates. With regard to RQ 1, the results revealed that the TG students had statistically higher scores on the dimensions of WTC with teacher and class ($F_{(1, 125)} = 12.743$, $p < .001$) and WTC with non-Chinese ($F_{(1, 125)} = 34.709$, $p < .001$). However, there was no significant between-subjects difference on WTC with friends ($F_{(1, 125)} = 0.929$, $p = .337$).

Table 1. Descriptive statistics of WTC mean scores

WTC	Pre-intervention		Post-intervention	
	TG	CG	TG	CG
With teacher and class	36.49	37.90	43.87	40.10
With friends	16.33	16.21	19.32	18.53
With non-Chinese	18.16	17.65	21.78	17.79

4.2. Change in interactional features

As for RQ 2, descriptive statistics showed that the TG students outperformed their CG counterparts on all the three measures of interactional features (Table 2). Moreover, the mean values of NfM for both groups demonstrated a perceptible downward trend over time, and the means of LRE for both groups showed an oscillating pattern. As for "recast," the TG students demonstrated no specific pattern, while their CG counterparts showed an evident decreasing trend.

Table 2. Between-subjects effects on NfM, LRE and recast

Measures	Unit	TG	CG	p	η^2_p
NfM	U2	0.011	0.006	.018	.044
	U4	0.014	0.002		
	U6	0.005	0.002		
	U8	0.003	0.001		
LRE	U2	0.110	0.081	.115	.020
	U4	0.178	0.136		
	U6	0.153	0.112		
	U8	0.164	0.163		
recast	U2	0.005	0.005	.176	.015
	U4	0.003	0.002		
	U6	0.004	0.001		
	U8	0.007	0.000		

Note. $\alpha < .05$.

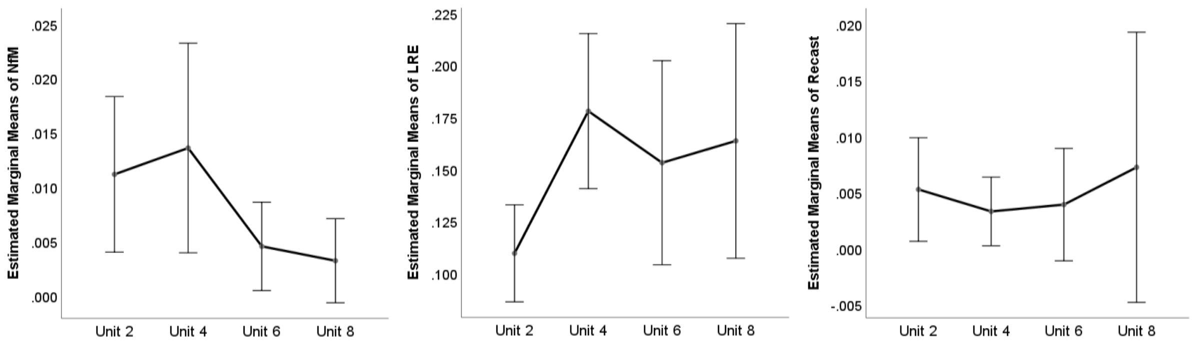
Levene's tests showed that the equality of error variances was not assumed on all the measures across the four units. However, a violation of this assumption is not considered an issue with roughly equivalent sample sizes on condition that the ratio of the largest sample size to smallest sample size is less than the threshold of 1.5 (Pituch & Stevens, 2016). The main effect of the between-subjects factor (i.e., group) on the mean values of NfM over time was statistically significant ($F_{(1, 125)} = 5.754$, $p = .018 < .05$) with a small to medium effect size ($\eta^2_p = 0.044 > 0.01$). In other words, the TG students generated more NfM per AS-unit than their CG counterparts. However, the TG students did not outperform their counterparts in the CG on LRE ($p = .115 > .05$) or recast ($p = .176 >$

.05) significantly, even though descriptive statistics showed that most of the TG mean scores appeared to be greater than the ones of the CG students.

4.3. Development trajectories of interactional features

In response to RQ 3, this study investigated the development trajectories of the TG students' interactional features. The results of Mauchly's test of sphericity showed that sphericity was not assumed (p < .001) for the three measures. The Greenhouse-Geisser epsilon for LRE was 0.783, greater than 0.75. Conversely, the Greenhouse-Geisser epsilon for NfM and recast was 0.724 and 0.552, respectively, both less than 0.75. Therefore, the Huynh-Feldt adjustment with the univariate tests was used for LRE, while the Greenhouse-Geisser adjustment was used for NfM and recasts. The tests of within-subjects effects revealed that the main effect of the within-subjects factor (i.e., time) was not statistically significant for NfM ($F_{(2.171, 143.302)} = 1.075$, p = .361 > .05), LRE ($F_{(2.478, 163.564)} = 0.110$, p = .930 > .05) or recast ($F_{(1.655, 109.213)} = 0.299$, p = .700 > .05), sphericity not assumed. In conclusion, the time factor did not lead to any statistically significant effect on any of the interactional features in the TG. Figure 4 illustrates the development trajectories of the three measures of the TG students' interactional features. Graphically, no particular pattern could be identified with regard to the effect of using ASR-based technology on the development of students' interactional features.

Figure 4. Development of trajectories of TG students' NfM, LRE and recast



Covariates appearing in the model are evaluated at the following values: placement = 82.25
Error bars: 95% CI

5. Discussion

5.1. Effect of ASR-based technology on Chinese EFL learners' WTC

The results of this study showed that owing to the use of the ASR-based technology in pre-class self-learning, the TG students' WTC with teacher and class and with non-Chinese was significantly improved. We argue that the ASR-enhanced practice changes the general notion pre-class self-study in a flipped classroom which is mainly passive absorption of knowledge (e.g., video-based self-learning). It allows the flipped EFL pedagogy to include an active component of interacting with ASR technology. Pedagogically, our approach enriches the current design for flipped classroom, especially for language-based flipped classroom approach.

Extant studies revealed that the use of the ASR-based technology could credibly simulate understanding by a native speaker, allowing learners to bridge the intelligibility gap and develop a sense of what successful interpersonal communication entails (Mroz, 2018). Moreover, as the ASR technology can immediately demonstrate the consequence of users' speech input, learners reported positive attitudes towards the use of the ASR-based technology for learning to speak (Ahn & Lee, 2016). Therefore, the ASR-based oral training is likely to establish a more friendly environment for learners to practice at their own pace (Wang & Young, 2014). As such, integrating ASR technology into FL oral training may contribute to lowering FL learners' anxiety and increasing their perceived competence, which helps foster learners' WTC in the target language (MacIntyre, 2007; MacIntyre, 2020).

Conversely, no significant difference in the participants' WTC with friends was observed. Attending an EFL class or talking with non-Chinese are specific occasions where the Chinese students must use English for communication. In contrast, talking with their friends in everyday life is a different occasion where the students do not necessarily have to speak in English. Since Asian societies are mostly monolingual and homogeneous, a great majority of Asian students do not speak English in daily life. In other words, on most occasions, they will only initiate daily conversations in their first language. Therefore, we speculate that the effects of the ASR-based technology on learners' WTC in oral English may only confine itself to occasions where the students must speak English and could not override the influence of the Chinese culture on interpersonal communication. But on a general basis, it is concluded that the use of ASR-based technology has positive impact on Chinese learners' WTC in oral English.

Recent research in other cultural contexts also obtained positive effects of technology-enhanced EFL learning on students' WTC (e.g., Lee & Sylvén, 2021), indicating that EFL teachers in the Chinese context should encourage students to use ASR technology for oral practice as frequently as possible in out-of-class learning. Accordingly, EFL teachers in monolingual societies such as the Chinese context should make better use of the ASR-based applications to support students' self-study of English. For example, EFL course developers and teachers should design more ASR-based oral practice for students' self-learning, which in turn may enhance their WTC and likely improve English oracy. On the other hand, policymakers and researchers in the Chinese EFL context should acknowledge that the integration of the ASR-based technology may not improve students' WTC in English with their friends. The reasons for that may be deeply rooted in the cultural impact and educational context, which may need to be further explored in future studies.

5.2. Effect of ASR-based technology on Chinese EFL learners' interactional features

Although the participants' WTC was enhanced through the adoption of the ASR-based technology, no significant improvement was observed in their in-class peer interaction, and the development trajectories of the TG students' interactional features showed no specific pattern. Such findings may indicate that the use of the ASR-based technology may only exert a limited effect on the participants' in-class peer interaction.

Peer interaction is a group-based and social behavior that involves at least two interlocutors in a communication. The dynamics among those interlocutors may be influenced by a range of factors from language competence, individual learning style to interpersonal relationship and local culture (MacIntyre, 2020). Inconsistent with the findings in previous studies where WTC was labelled as the decisive step to initiate a communication in the target language (MacIntyre & Doucette, 2010), the current study revealed that the improvement in WTC might not necessarily lead to more interactional features in peer interaction in the Chinese EFL classrooms. In other words, WTC may not function as a direct indicator of how well the Chinese students are engaged in in-class peer interaction in an EFL classroom. Future studies need to investigate the relationship between Chinese EFL learners' WTC and their actual behaviors of peer interaction in class. Particularly, since the Chinese educational practice and local culture seem to have a more direct effect on the students' peer interaction than WTC, follow-up research is expected to determine what contextual factors may mediate the effect of WTC on peer interaction in the Chinese EFL context.

Given the students' inexperience in learning spoken English, the integration of the ASR-based technology into their pre-class self-learning may not facilitate the improvement of their English oracy within such a short period of 14 weeks. Besides, since the input of the ASR-based tasks was mainly self-produced by the students, most of whom were not as competent in oral English, both the quality and the quantity of such input might be inadequate and limited. Therefore, the TG students using ASR-based technology might need to take longer to attain improvements similar to those typically observed in naturalistic contexts (Hanzawa, 2021). In EFL pedagogy, the ASR-based pre-class tasks may need re-designing to better prepare the students for in-class peer interaction. Although it is a dictation ASR-based application which does not provide intended feedback as Google Assistant, iFlyRec may be used together with Google Assistant in future research to see whether the two kinds of ASR technologies could jointly improve Chinese EFL learners' peer interaction and further enhance iCALL in the Chinese EFL context. Moreover, to understand the holistic process of the technology use, future studies may study how the students use the ASR-based applications for oral practice and whether there are unforeseen usage behaviors on the learners' side.

Social context may also act as a critical factor in classroom-based interaction among Chinese EFL learners. In the Chinese context, communicating in English and providing corrective feedback to peers can easily be interpreted as an act of showing off, which may make their peers lose face. This was corroborated by Tomita and Spada's (2013) study among Japanese EFL learners, whose culture is deeply influenced by the Chinese culture. Conversely, the non-significant effect may also result from the test-oriented educational practice in the Chinese context. The Chinese students tend to focus on "knowledge on test papers," meaning what is tested in examinations or even quizzes will be considered as "useful knowledge" by them simply because such knowledge can score and make them rank higher. There is no doubt that the prevalent examination-oriented learning culture substantially impacts teachers' teaching and students' learning behavior (Hu & West, 2014). Most Chinese students in high school learned English as an academic subject, of which the examination score served as a "ladder" for them to enter university. Since speaking is not tested in most middle schools in China, the students spent almost all their time on reading and writing practice (Liu & Chen, 2018). This kind of learning experience bears a considerable gap in tertiary EFL education. The gap between students' perceived mastery of exam-oriented English and the competence of authentic language use may constitute a major obstacle for most of the students to interact with each other orally in class.

In view of the above, it is worth mentioning that the use of educational technologies in EFL classrooms should be given full consideration of the Chinese educational context in practice. While the advantages of the ASR-based technology seem evidently conducive to developing students' English oracy in extant studies, as found in this study, its effect on facilitating task-based peer interaction in class may not be as expected. The discrepancy may be probably attributable to students' unforeseen and even distorted use of the ASR technology, which is beyond the scope of the current study. Accordingly, it is recommended that EFL teachers try to design and implement learning tasks that fit into the Chinese educational context. As Chinese university students are more likely to remain reticent in class (Tang et al., 2020), the tasks in Chinese EFL classrooms should be deliberately designed to have a meaningful connection with the students' lives. Besides, authenticity is regarded as a critical feature in task design. When the tasks cannot resemble real-life language use, the students may be less inclined to get involved. Meanwhile, to reduce Chinese students' speaking anxiety, teachers need to elaborate scaffolded tasks (Evers & Chen, 2020) by providing relevant organizational structures and flexible templates that best align to students' learner preparedness in a flipped EFL classroom. The scaffolded tasks should be contingent on students' task readiness, which requires skillful assessment for instructor to peg the task at the right level of zone of proximal development to optimize sustained engagement for the flipped approach. More importantly, helping Chinese EFL students to develop an English user identity may lead them to engage in positive peer interactions.

6. Conclusion, implications and limitations

The present study found that integrating the ASR-based technology into pre-class self-learning in a flipped EFL classroom could improve Chinese students' WTC with teacher and class as well as with non-Chinese. Contrarily, no significant difference was noticed in students' WTC with friends. With respect to the task-based peer interaction, despite the significant between-subjects difference in NfM, this study revealed that the Chinese students' interactional features in the EFL classroom did not improve significantly over time, indicating that the effects of the ASR-based technology on Chinese EFL learners' interactional features seemed to be quite limited.

The findings from the current study may prompt ASR proponents in EFL learning to revisit its effectiveness in the Chinese culture and the overriding impact of local culture on peer interaction in English class. Although task-based peer interactions are regarded as learning opportunities in EFL classrooms, the interactional features of the Chinese students' in-class peer interaction may not occur as expected. Therefore, course developers and practitioners need to take the cultural influence into serious consideration when designing EFL learning activities for Chinese learners. Chinese EFL teachers may need to come up with learning and teaching strategies to mitigate the effects of the local culture. Orienting the students to the benefits of task-based peer interactions in EFL classrooms is also necessary.

One major limitation of this study is the sample size. The participants all came from a single university in China, which might lead to a representativeness issue for the generalization of the findings. Another limitation of the present study is that, owing to the university rules, we were not allowed to collect data in person in the classroom. Therefore, the participants recorded their own task performance in class, which may have resulted in the loss of some raw data. Moreover, longer term research is needed to examine whether the novelty effect may contribute to the positive effect of the ASR-based technology on students' WTC revealed in the present study.

Acknowledgement

Our special thanks go to the Chinese University of Hong Kong Stanley Ho Big Data Decision Analytics Research Centre and the Chinese University of Hong Kong Teaching Development and Learning Enhancement Grant awarded to Professor Helen Meng for supporting this interdisciplinary research.

References

Ahn, T. Y., & Lee, S. M. M. (2016). User experience of a mobile speaking application with automatic speech recognition for EFL learning. *British Journal of Educational Technology, 47*(4), 778-786. https://doi.org/10.1111/bjet.12354

Anderson, L. W., & Krathwohl, D. R. (2001). *A Taxonomy for learning, teaching, and assessing: A Revision of Bloom's taxonomy of educational objectives.* Longman.

Clark, R., & Gieve, S. N. (2006). On the discursive construction of "the Chinese learner." *Language, Culture and Curriculum, 19*(1), 54-73. https://doi.org/10.1080/07908310608668754

Durkin, K. (2011). Adapting to Western norms of critical argumentation and debate. In L. Jin, & M. Cortazzi (Eds.), *Researching Chinese Learners: Skills, Perceptions, and Intercultural Adaptations* (pp. 274-291). Palgrave Macmillan.

Evers, K., & Chen, S. (2020). Effects of an automatic speech recognition system with peer feedback on pronunciation instruction for adults. *Computer Assisted Language Learning.* https://doi.org/10.1080/09588221.2020.1839504

Foster, P., Tonkyn, A., & Wigglesworth, G. (2000). Measuring spoken language: A Unit for all reasons. *Applied Linguistics, 21*(3), 354-375. https://doi.org/10.1093/applin/21.3.354

Gao, X. S. (2008). You had to work hard 'cause you didn't know whether you were going to wear shoes or straw sandals! *Journal of Language Identity and Education, 7*(3-4), 169-187. https://doi.org/10.1080/15348450802237798

Gass, S., Mackey, A., & Ross-Feldman, L. (2011). Task-based interactions in classroom and laboratory settings. *Language Learning, 61*(S1), 189-220. https://doi.org/10.1111/j.1467-9922.2011.00646.x

Hanzawa, K. (2021). Development of second language speech fluency in foreign language classrooms: A Longitudinal study. *Language Teaching Research.* https://doi.org/10.1177/13621688211008693

Hayes, A. F., & Krippendorff, K. (2007). Answering the call for a standard reliability measure for coding data. *Communication Methods and Measures, 1*(1), 77-89. https://doi.org/10.1080/19312450709336664

Hincks, R. (2015). Technology and leaning pronunciation. In M. Reed, & J. Levis (Eds.), *The Handbook of English Pronunciation* (pp. 505-519). John Wiley & Sons, Inc.

Hu, B., & West, A. (2014). Exam-oriented education and implementation of education policy for migrant children in urban China. *Educational Studies, 41*(3), 249-267. https://doi.org/10.1080/03055698.2014.977780

Jiang, M. Y. C., Jong, M. S. Y., Lau, W. W. F., Chai, C. S., & Wu, N. (2021a). Using automatic speech recognition technology to enhance EFL learners' oral language complexity in a flipped classroom. *Australasian Journal of Educational Technology, 37*(2), 110-131. https://doi.org/10.14742/ajet.6798

Jiang, M. Y. C., Jong, M. S. Y., Lau, W. W. F., Meng, Y. L., Chai, C. S., & Chen, M. Y. (2021b). Validating the general extended technology acceptance model for e-learning: Evidence from an online EFL course amid COVID-19. *Frontiers in Psychology, 12*, 671615. https://doi.org/10.3389/fpsyg.2021.671615

Jiang, M. Y. C., Jong, M. S. Y., Lau, W. W. F., Chai, C. S., Liu, K. S. X., & Park, M. (2022a). A scoping review on flipped classroom approach in language education: Challenges, implications and an interaction model. *Computer Assisted Language Learning, 35*(5-6), 1218-1249. https://doi.org/10.1080/09588221.2020.1789171

Jiang, M. Y. C., Jong M. S. Y., Lau, W. W. F., Chai, C. S., & Wu, N. (2022b). Exploring the effects of automatic speech recognition technology on oral accuracy and fluency in a flipped classroom. *Journal of Computers Assisted Learning*. https://doi.org/10.1111/jcal.12732

Jong, M. S. Y. (2017). Empowering students in the process of social inquiry learning through flipping the classroom. *Educational Technology & Society, 20*(1), 306-322.

Jong, M. S. Y. (2019). To flip or not to flip: Social science faculty members' concerns about flipping the classroom. *Journal of Computing in Higher Education, 31*(2), 391-407. https://doi.org/10.1007/s12528-019-09217-y

Jong, M. S. Y., Chen, G. W., Tam, V., & Chai, C. S. (2019). Adoption of flipped learning in social humanities education: The FIBER experience in secondary schools. *Interactive Learning Environments, 27*(8), 1222-1238. https://doi.org/10.1080/10494820.2018.1561473

Kim, I. S. (2006). Automatic speech recognition: Reliability and pedagogical implications for teaching pronunciation. *Educational Technology & Society, 9*(1), 322-334.

Lee, J. C. K., & Yin. H. B. (2011). Teachers' emotions and professional identity in curriculum reform: A Chinese perspective. *Journal of Educational Change, 12*(1), 25-46. https://doi.org/10.1007/s10833-010-9149-3

Lee, J. S., & Sylvén, L. K. (2021). The Role of informal digital learning of English in Korean and Swedish EFL learners' communication behaviour. *British Journal of Educational Technology, 52*, 1279-1296. https://doi.org/10.1111/bjet.13082

Levis, J. (2007). Computer technology in teaching and researching pronunciation. *Annual Review of Applied Linguistics, 27*, 184-202. https://doi.org/10.1017/S0267190508070098

Liu, S., & Chen, Y. J. (2018). A Practical exploration on NMET(Shanghai)-based English listening and speaking teaching. *Technology Enhanced Foreign Language Education, 5*, 30-34. https://doi.org/CNKI:SUN:WYDH.0.2018-05-006.

Loewen, S., & Sato, M. (2018). Interaction and instructed second language acquisition. *Language Teaching, 51*(3), 285–329. https://doi.org/10.1017/S0261444818000125

Long, M. H. (1981). Input, interaction, and second language acquisition. *Annals of the New York Academy of Sciences, 379*(1), 259-278. https://doi.org/10.1111/j.1749-6632.1981.tb42014.x

Long, M. H. (2015). *Second language acquisition and task-based language teaching.* Wiley.

Lou, N. M., & Noels, K. A. (2021). Western and heritage cultural internalizations predict EFL students' language motivation and confidence. *International Journal of Bilingual Education and Bilingualism, 24*(5), 636-650. https://doi.org/10.1080/13670050.2018.1508277

MacIntyre, P. D. (2007). Willingness to communicate in the second language: Understanding the decision to speak as a volitional process. *Modern Language Journal, 91*, 564-576. https://doi.org/10.1111/j.1540-4781.2007.00623.x

MacIntyre, P. D. (2020). Expanding the theoretical base for the dynamics of willingness to communicate. *Studies in Second Language Learning and Teaching, 10*(1), 111-131. https://doi.org/10.14746/ssllt.2020.10.1.6

MacIntyre, P. D., & Doucette, J. (2010). Willingness to communicate and action control. *System, 38*, 161-171. https://doi.org/10.1016/j.system.2009.12.013

Mackey, A., & Goo, J. (2007). Interaction research in SLA: A Meta-analysis and research synthesis. In A. Mackey (Ed.), *Conversational Interaction in Second Language Acquisition* (pp. 407-452). Oxford University Press.

McCrocklin, S. (2016). Pronunciation learner autonomy: The Potential of automatic speech recognition. *System*, *57*, 25-42. https://doi.org/10.1016/j.system.2015.12.013

McCrocklin, S. (2019). ASR-based dictation practice for second language pronunciation improvement. *Journal of Second Language Pronunciation, 5*(1), 98-118. https://doi.org/10.1075/jslp.16034.mcc

Mroz, A. (2018). Seeing how people hear you: French learners experiencing intelligibility through automatic speech recognition. *Foreign Language Annals, 51*(3), 617-637. https://doi.org/10.1111/flan.12348

Nicholas, H., Lightbown, P. M., & Spada, N. (2001). Recasts as feedback to language learners. *Language Learning, 51*, 719-758. https://doi.org/10.1111/0023-8333.00172

Norris, J. M., & Ortega, L. (2009). Towards an organic approach to investigating CAF in instructed SLA: The Case of complexity. *Applied Linguistics, 30*(4), 555-578. https://doi.org/10.1093/applin/amp044

Pawlak, M., & Mystkowska-Wiertelak, A. (2015). Investigating the dynamic nature of L2 willingness to communicate. *System*, *50*, 1-9. https://doi.org/10.1016/j.system.2015.02.001

Philp, J., Adams, R., & Iwashita, N. (2014). *Peer interaction and second language learning.* Routledge.

Pica, T. (1994). Questions from the language classroom: Research perspectives. *TESOL Quarterly*, *28*, 49-79. https://doi.org/10.2307/3587198

Pituch, K. A., & Stevens, J. P. (2016). *Applied multivariate statistics for the social sciences* (6th ed.). Routledge.

Rassaei, E. (2022). Recasts during mobile-mediated audio and video interactions: Learners' responses, their interpretations, and the development of English articles. *Computers Assisted Language Learning.* https://doi.org/10.1080/09588221.2019.1671461

Robinson, P., & Gilabert, R. (2007). Task complexity, the cognition hypothesis, and second language learning and performance. *International Review of Applied Linguistics, 45*(3), 161-176. https://doi.org/10.1515/IRAL.2007.007

Ryan, R. M., & Deci, E. L. (2020). Intrinsic and extrinsic motivation from a self-determination theory perspective: Definitions, theory, practices, and future directions. *Contemporary Educational Psychology, 61*, 101860. https://doi.org/10.1016/j.cedpsych.2020.101860

Sato, R. (2020). Examining fluctuations in the WTC of Japanese EFL speakers: Language proficiency, affective and conditional factors. *Language Teaching Research*. https://doi.org/10.1177/1362168820977825

Sun, Z. R., & Xie, K. (2020). How do students prepare in the pre-class setting of a flipped undergraduate math course? A Latent profile analysis of learning behavior and the impact of achievement goals. *The Internet and Higher Education, 46*, 100731. https://doi.org/10.1016/j.iheduc.2020.100731

Swain, M., & Lapkin, S. (1998). Interaction and second language learning: Two adolescent French immersion students working together. *Modern Language Journal, 82*, 320-337. https://doi.org/10.1111/j.1540-4781.1998.tb01209.x

Tai, T. Y., & Chen, H. H. J. (2020). The Impact of Google Assistant on adolescent EFL learners' willingness to communicate. *Interactive Learning Environments.* https://doi.org/10.1080/10494820.2020.1841801

Tang, X. W., Wang, Y., & Wong, D. (2020). Learning to be silent: Examining Chinese elementary students' stories about why they do not speak in class. *Language, Culture and Curriculum, 33*(4), 384-401. https://doi.org/10.1080/07908318.2020.1715999

Tavakoli, E., & Davoudi, M. (2017). Willingness to communicate orally: The Case of Iranian EFL learners. *Journal of Psycholinguistic Research, 46*, 1509-1527. https://doi.org/10.1007/s10936-017-9504-0

Tomita, Y., & Spada, N. (2013). Form-focused instruction and learner investment in L2 communication. *The Modern Language Journal, 97*(3), 591-610. https://doi.org/10.1111/j.1540-4781.2013.12031.x

Tran, T. T. (2013). Is the learning approach of students from the Confucian heritage culture problematic? *Educational Research for Policy and Practice, 12*(1), 57-65. https://doi.org/10.1007/s10671-012-9131-3

Wang, C. C., Tseng, W. T., Chen, Y. L., & Cheng, H. F. (2019). Classroom interactions in the target language: Learners' perceptions, willingness to communicate, and communication behavior. *Asia-Pacific Education Researcher, 29*(5), 393-404. https://10.1007/s40299-019-00492-y

Wang, Y. H., & Young, S. S. C. (2014). A Study of the design and implementation of the ASR-based iCASL system with corrective feedback to facilitate English learning. *Educational Technology & Society, 17*(2), 219-233.

Wilson, M. L., & Zietz, J. (2004). Systematic bias in student self-reported data. *Journal for Economic Educators, 4*(4), 13-19.

Wu, Q. (2015). Re-examining the 'Chinese learner': A Case study of mainland Chinese students' learning experiences at British universities. *Higher Education, 70*(4), 753-766. https://doi.org/10.1007/s10734-015-9865-y

Xu, J. F., & Kou, J. N. (2018). Group interaction strategies and students' oral performance in Chinese EFL classrooms. *TESOL Quarterly, 52*(1), 198-209. https://doi.org/10.1002/tesq.398

Yum, J. O. (2007). Confucianism and communication: Jen, Li and Ubuntu. *China Media Research, 3*(4), 15-22. http://chinamediaresearch.net/index.php/back-issues?id=34

Zare, M., Shooshtari, Z. G., & Jalilifar, A. (2020). The Interplay of oral corrective feedback and L2 willingness to communicate across proficiency levels. *Language Teaching Research, 26*(6). https://doi.org/10.1177/1362168820928967

Zhang, J. Y., Beckmann, N., & Beckmann, J. F. (2020). More than meets the ear: Individual differences in trait and state willingness to communicate as predictors of language learning performance in a Chinese EFL context. *Language Teaching Research.* https://doi.org/10.1177/1362168820951931

Appendix A. Items in the WTC questionnaire

1. WTC with teacher and class (13 items)

WTCTC1. When the teacher's instruction for a task in classroom is not clear, I feel relaxed to ask in English from the teacher.
WTCTC2. For speaking in English in the class, I like to wait for my own turn and wait for the teacher to ask me to speak.
WTCTC3. When the teacher asks a question from all the students (for instance when asking the class opinion), I am willing to immediately answer in English.
WTCTC4. I am willing to participate in classroom discussions in English voluntarily.
WTCTC5. In classroom discussions, I am more willing to wait for the time the teacher asks me in person.
WTCTC6. I am not among the students who voluntarily start speaking in English in the classroom.
WTCTC7. I am willing to take part in classroom group works and speak in English.
WTCTC8. After classroom group discussions/speaking tasks, I am willing to be a volunteer to report the results.
WTCTC9. When I start speaking in English in the class in front of all my friends, I lose my confidence and concentration.
WTCTC10. I am willing to express my thoughts, opinions and even emotions in English in the class to all.
WTCTC11. I feel relaxed to share my opinions and even emotions with my teammates in English.
WTCTC12. I volunteer to orally present the lessons or talk about a topic in English to all the students.
WTCTC13. I prefer to keep silent in the class because speaking in English makes me agitated.

2. WTC with friends (7 items)

WTCF1. I am willing to speak in English with my classmates before the class begins.
WTCF2. If I have questions about the assigned homework for the next class, I prefer to ask in English from

the students next to me.

WTCF3. When the teacher's instruction for a task in classroom is not clear, I feel relaxed to ask in English from my friends sitting next to me.

WTCF4. I have the desire to communicate in English with my classmates on the first day of the class.

WTCF5. I like to use every opportunity in the class (like the break time or the spare time of the group works) to speak in English with my friends.

WTCF6. I have the desire to speak in English with my former classmates or teachers outside of the class at the language school.

WTCF7. For group speaking tasks, I am more willing to have students in my group who let me speak more.

3. WTC with non-Chinese (7 items)

WTCNS1. When I see a tourist in street, hotel, restaurant, or park, I try to find an excuse so that I can approach them and speak with them in English.

WTCNS2. If I travel to an English-speaking country, I feel relaxed to approach the people in the street, parks etc. and start communicating in English.

WTCNS3. I am willing to communicate in English with people who speak English as their first language.

WTCNS4. If I encounter people who can speak English as fluently as a native speaker, I will easily start speaking with them despite not knowing them in advance.

WTCNS5. I like to learn speaking English so that I can communicate orally with the English speakers.

WTCNS6. I am willing to be a tour guide for some days (even free of charge) so that I can communicate in English with the native speakers who have come to visit my city (such as Australians, Canadians).

WTCNS7. I am willing to be a tour guide for some days (even free of charge) so that I can communicate in English with the non- native speakers who have come to visit my city (such as Pakistanis, Japanese).

Appendix B. Definitions of the measures of interactional features (Gass et al., 2011)

Measures	Definitions	Working examples*
NfM	An attempt to overcome comprehension problems, such as confirmation checks, clarification requests and comprehension checks.	**Student 1:** … and I don't, I don't like to eat some French, eh, French things, like French fries in the morning because it isn't very healthy. **Student 2:** French fries? ***(Confirmation check)*** **Student 1:** Yeah. But I think the healthy diet is to, eh, keep your meals in order, … *[Class 1, Group 5, Unit Task 2: Food and diet]*
LRE	Any part of a dialogue in which students talk about the language they are producing, question their language use or other- or self-correct, such as corrective feedback and self-correction.	**Student 1:** Many people like west, eh, western ***(Self-correction)*** food, but I think Chinese food is very delicious, like, eh, noo…noodles. **Student 2:** Rice. *[Class 3, Group 2, Unit Task 2: Food and diet]*
recast	Correct restatement of a learner's incorrectly formed utterance.	**Student 1:** Oh, my parents like some, some place where it is quite, quite. **Student 2:** Quite? What the meaning? **Student 1:** Quite, such as quite, peace. **Student 2:** Quiet! ***(Recast)*** **Student 1:** Quiet. Okay, fine, quiet. *[Class 4, Group 1, Unit Task 6: Dream trip]*

Note. *The examples were directly obtained from the data collected in this study.

Kohnke, L., Foung, D., & Chen, J. (2023). Cluster Analysis of Hong Kong Students' Self-Regulated Learning in Contextualized Multimodal Language Learning. *Educational Technology & Society*, *26*(3), 53-68. https://doi.org/10.30191/ETS.202307_26(3).0005

Cluster Analysis of Hong Kong Students' Self-Regulated Learning in Contextualized Multimodal Language Learning

Lucas Kohnke[1], Dennis Foung[2*] and Julia Chen[3]

[1]The Education University of Hong Kong, China // [2]The University of British Columbia, Canada // [3]The Hong Kong Polytechnic University, China // lmakohnke@eduhk.hk // dennis.foung@ubc.ca // julia.chen@polyu.edu.hk

[*]Corresponding author

ABSTRACT: This study investigated how English learners complete multimodal formative quizzes. Participants included 17,950 students enrolled in a mandatory English for Academic Purposes course at a university in Hong Kong. We retrieved data from Blackboard, a learning management system, and conducted a two-step cluster analysis to examine student self-regulated learning (SRL) profiles with the quizzes. We first identified five clusters of learners with distinctively different self-regulated learning patterns. Then, we performed a multivariate analysis of variance (MANOVA) to further explore their differences in SRL, in terms of start day, days started before deadline, differences in scores between first and last attempt, and scores in language learning activities. Our findings echoed those of previous studies on the relationship between self-regulated learning and academic success. This research enables us to better understand the needs of EAP students in Hong Kong.

Keywords: Cluster analysis, English for Academic Purposes, Multimodal, Formative, Quizzes

1. Introduction

Digital multimodal language learning (e.g., audio, videos, cartoons, infographics) is becoming an integral component of English language teaching (ELT) (Jiang & Ren, 2021; Kohnke & Jarvis, 2022). The benefits of multimodality in ELT include heightened semiotic awareness, multiple modes of input, and enhanced communicative competence (Hafner & Miller, 2011; Shin & Cimasko, 2008). In English for Academic Purposes (EAP) courses in Hong Kong, multimodal language learning is used to facilitate second-language acquisition (Hafner & Pun, 2020; Kohnke et al., 2021). By aligning technology, second-language pedagogy (Chapelle & Sauro, 2017), and multimodality (Yeh, 2018), teachers can develop authentic, engaging formative assessments that foster independent language learning (Park et al., 2016).

Formative assessments such as quizzes are integral to monitoring the knowledge and skills of second-language learners (Gardner, 2012; Hinkelman, 2018). Online formative assessment is defined as the use of technological tools to support the process of "gathering and analysing information about student learning by teachers…" (Pachler et al., 2010, p. 716). In higher education, this is usually done by designing useful online activities that can provide feedback to learners on their learning progress. Using automated multimodal quizzes in an EAP program allows students to self-assess while simultaneously requiring them to employ a variety of online learning strategies (Wandler & Imbriale, 2017). Self-regulation skills are a critical variable for success at language learning (Dörnyei, 2005; Tseng et al., 2006). Studies have reported the importance of self-regulation in blended learning environments (Artino, 2007; Broadbent, 2017), though none have focused on multiple semesters of a large EAP course.

Although previous studies have reported on learners' behaviors regarding in-class quizzes (Ross et al., 2018) and multimodality (Kohnke et al., 2022), EAP learners' SRL with multimodal quizzes remains underexplored. Moreover, learning-analytics-based studies that aim to generate actionable insights have not been common in EAP SRL research. Understanding students' self-regulation profiles is important for better developing course policies to target specific types of SRL profiles. Therefore, this study examined students' SRL regarding multimodal quizzes. Using a two-step cluster analysis, we explored the self-regulation profiles of 17,950 students who had taken an EAP course between 2012 and 2019, focusing on how they accessed the online multimodal language-learning content.

2. Literature review

2.1. Multimodality in language learning

One common definition of multimodality is “the use of several semiotic modes in the design of a semiotic product or event… in which these modes are combined… [to] reinforce each other…and fulfil complementary role[s]” (Kress & van Leeuwen, 2001, p. 20). This definition suggests that multimodality combines multiple input methods, such as text, sound, and/or video. The belief that multimodality benefits learning originated from the insight, gained from dual coding theory, that learning can be better facilitated if the information is processed in both spoken and written modes (Paivio, 1986). Earlier research (before the advent of computer-assisted technology) on multimodality in language learning focused on how multimodality can expose learners to diverse ways of communicating and making meaning (Hampel & Hauck, 2006), involving the use of non-computer-assisted multimodality, such as visual, verbal, and other means (Kendrick et al., 2006). However, multimodal language learning research within a non-technological context is still prevalent, using storybooks with pictures and audio input (Tragant & Pellicer-Sánchez, 2019) or videos with text subtitles (Peters, 2019; Pujadas & Muñoz, 2019). This line of research clearly illustrates the benefits of multimodality well before the era of computer-assisted language learning.

Since the advent of research on computer-assisted language learning, research on multimodality is now equally, if not more, interested in how combinations of videos, audios, texts, and online interactive resources can enhance language learning in a computer-assisted environment. For example, Marcel (2020) explored the use of augmented reality (AR) and virtual reality (VR) as a multimodal approach to language learning. They found that learners gained vocabulary through a contextualized multimodality experience. White et al. (2021) explored the use of videoconferencing tools in language learning and found that besides the video and audio inputs provided by the tools, the photo-sharing function created more possibilities for language learning, as learners were not only stimulated by the audio and video footage of the teacher but also by photos shared during the session. Other studies have been conducted on multimodal feedback on language learning, such as those by Wilkie and Liefeith (2020) and Martin (2020). These studies all demonstrate that multimodal resources offering sound, image, text, and animation yield opportunities for effective and dynamic learning.

While these studies establish the effectiveness of multimodality in language learning, they explore the use of multimodality over shorter periods (e.g., a semester and a year); in fact, many studies have examined the effectiveness of multimodal interventions within a course for no more than a year (e.g., Marcel, 2020; Mauricio & Genuino, 2020; White et al., 2021; Wilkie & Liefeith, 2020). This suggests a need for research on multimodal learning over a more prolonged period.

2.2. Multimodality in higher education

In higher education, multimodality is often introduced with various blended-learning input methods through learning management systems (LMS). An LMS provides avenues for multimodal blended learning using tools such as videos, pages, discussion forums, and quizzes (Cole et al., 2021; Coskuncay & Ozkan, 2013). These modalities complement each other in enhancing the learning experience. Studies have found that, overall, students learning in blended classes perform better than those in face-to-face-only classes (Garrison & Vaughan, 2013; Ross & Gage, 2006; Porter et al., 2016; Owston & York, 2018). Among multimodal tools, recent studies have illustrated the positive effects of online quizzes that include multimodal elements (Cook & Babon, 2017; Gamage et al., 2019). It is therefore crucial to examine the potential of these multimodal elements in online quizzes so that teachers can determine the type of multimodal activities that will best support the pedagogical process (Lamy, 2012).

Research on online quizzes suggest that they allow the monitoring of progress and provide timely feedback to support learning (Nicole & Macfarlane-Dick, 2006). Automated online quizzes can be particularly helpful for second-language learners, allowing them to self-assess and take action to address weaknesses (independently or with the help of instructors or peers). While most students consider online quizzes non-threatening (Gardner, 2012), some will not complete them unless they believe the quizzes will make a substantial difference in their ability to succeed in a course. Accordingly, teachers tend to encourage completion by assigning a small percentage of the course grade to each quiz (Padilla-Walker, 2006). To maximize second-language learning, learners need to attempt the quizzes repeatedly. Such quizzes have been found to increase student enthusiasm, achievement levels, and self-regulation (McLaughlin & Yan, 2017). However, research using clustering with multimodal online formative quizzes has been limited in the EAP context.

2.3. Online self-regulated learning

According to Zimmerman (1990), all learners self-regulate to a certain degree. Self-regulated learning (SRL), which entails being systematic in one's learning (Zimmerman & Schunk, 2011), is an important indicator of effectiveness in a face-to-face learning environment (Boekaerts, 1999). Self-regulation involves being active and goal-directed and displaying self-control, motivation, and cognition in performing academic tasks (Pintrich, 1995). These traits are equally important in the online learning environment, where learners have a high degree of autonomy and little teacher presence (Lehmann et al., 2014). Previous studies have found SRL to be an important predictor of learner achievement (Broadbent & Poon, 2015; Kuo et al., 2013) and success in online studies (Bol & Garner, 2001; Cho & Heron, 2015). Students exhibiting successful SRL will set learning goals, plan tasks, and monitor their progress even when facing academically challenging tasks (Broadbent & Poon, 2015; Cho & Cho, 2017). Confidence is also an important factor in SRL, as confident students participate online more strategically (Cho & Jonassen, 2009) and are more likely to set goals and to monitor and adjust their learning processes (Cho & Cho, 2017). A positive attitude is indispensable for engagement in SRL processes (Pintrich, 2004; Zimmerman & Schunk, 2011).

Furthermore, task value (i.e., the perceived value of starting and completing a task) greatly influences SRL. Learners who place a high task value on academic work set clear goals, monitor their learning systematically, and adopt strategies to accomplish their goals (Cho & Shen, 2013; Lawanto et al., 2014). Cho and Heron (2015) found that students who received a passing grade in an online course showed higher task value and motivation than non-passing students. Learners who are less skillful at SRL often fail to set learning goals and demonstrate low confidence in their learning and the learning process. They tend to blame their performance on the instructor or materials. Self-efficacy impacts task choice, effort, persistence, and achievement (Schunk & Pajares, 2002). Students with positive self-efficacy tend to perform better in online courses (Wang et al., 2013). This correlation between SRL and online academic success, which is supported by previous studies (Azevedo & Hadwin, 2005), indicates a need to provide SRL support to all students. To this end, Hill and Hannafin (2001) suggested four types of support: (i) help in prioritizing information, (ii) metacognitive support (e.g., asking questions that help students reflect), (iii) help with resources (e.g., assistance in locating appropriate learning tools), and (iv) multiple options for completing a task.

2.4. Person-centered approach to SRL

Due to the complex nature of SRL, scholars advocate a "person-centered" approach to SRL that explores whether there can be subgroups of learners with distinct SRL behaviors and whether these sub-groups differ in important external criteria (Broadbent & Fuller-Tyszkiewicz, 2018, p. 1437). Person-centered investigation of SRL provides useful insights to course designers on how specific strategies and course policies can be adopted by teachers to promote SRL. With the emergence of data available on online systems, more studies prefer the use of trace data to examine students' SRL, rather than self-reported SRL. One major challenge of using trace data is the measurement of SRL, because there is no single measure that can fully represent all SRL (Winne & Perry, 2000). Also, while some SRL indicators can be observed in the environment (e.g., performance), many SRL indicators can only be inferred (Winne & Perry, 2000). For example, trace data can show some students accessing materials earlier than other students, but one can only infer that the students who access materials early do so to plan their study. There are likely to be challenges in operationalizing SRL variables based on student behavioral traces.

To identify typologies in SRL / adopt a person-centered approach, most studies adopt cluster analysis, according to a recent review (Elsayed et al., 2019). Cluster analysis is an exploratory analysis that attempts to divide samples into groups so that the degree of association for variables within a group is minimal and for other groups is maximal (Antonenko et al., 2012). While some literature considers cluster analysis to be like factor analysis, cluster analysis can also be viewed as a way of visualizing different groups of samples in a large data set (Antonenko et al., 2012). Unlike traditional inferential statistics that requires testing of assumption, cluster analysis can be conducted based on the types of data. For example, k-mean clustering adopted in this study, is defined as non-hierarchical in nature and can take continuous or nominal data but the number of clusters needed to be determined (Antonenko et al., 2012). Recent SRL studies with cluster analysis usually begin by establishing the number of clusters using methods such as the Elbow methods (Yuan & Yang, 2019), followed by an examination of the clusters. The analysis is usually concluded by examining the differences between clusters regarding some external variables. See Ng et al. (2016) and Broadbent and Fuller-Tyszkiewicz (2018) for SRL studies and Guo et al. (2022) and Stenlund et al. (2018) for other educational studies. These studies, especially the SRL studies, can successfully identify and discuss clusters in terms of SRL and other external variables (e.g., course outcomes).

While these studies paint a vivid picture of how SRL profiles can enhance teachers' and course designers' understanding and enable them to develop targeted strategies for students, not enough person-centered SRL studies have been conducted in the EAP context, and more specifically with contextualized multimodal learning. Therefore, this study aims to examine the SRL profiles of students in completing contextualized multimodal quizzes.

3. Methodology

To understand and optimize learning, this study adopted a learning-analytics approach (Ferguson, 2012) to collect and analyze data on students' SRL behaviors with multimodal content. Data were retrieved from the LMS, and a two-step cluster analysis was conducted to identify student SRL profiles. Ethical clearance was obtained, and the learning data were retrieved after a formal data request was made as stipulated by the data governance framework of the research site.

3.1. Context

The study used data gathered from students taking a 13-week EAP course at a university in Hong Kong. The course was launched in 2012 when a 4-year undergraduate curriculum was introduced in Hong Kong. It is a mandatory English course in the undergraduate curriculum, with an annual enrolment of 2,500 to 3,000 students.

The course is standardized across cohorts in terms of course assessments, grading descriptors, and marker training procedures. While its teachers have flexibility in delivering class activities and may provide additional class materials, the course notes, assessments, grading criteria, descriptors, and multimodal quiz requirements were comparable across the cohorts and classes included in this study. The course is assessed through two essays and one presentation. Grading criteria include content development (e.g., in essays), organization (including source incorporation), language (including style), and referencing skills. These three assessment categories determine the overall course grade, which is reduced by a penalty if students fail to complete the multimodal quiz requirements.

3.2. Multimodal learning package

The multimodal learning package (MLP) is composed of numerous activities hosted on Blackboard, the university's LMS. In earlier cohorts (2012–2014), there were more than 15 activities each semester, but the activities were re-grouped to 13 from 2015 onwards. The activities cover four areas: academic style, genre knowledge, referencing, and academic presentation skills. Each activity contains multimodal content (e.g., videos, podcasts, reading, and infographics) and is followed by an online formative quiz with around 20 questions. As a formative assessment, students will know the correct answers of the quiz after submission so that they can know how well they did, i.e., their learning progress. The activities are designed to supplement the content taught in class (see the Appendix for details). For example, after the discussion of academic style during class in Week 1, students are expected to complete an activity on academic style as the "Session 1" activity.

The MLP was designed by experienced in-house teachers and was first piloted in 2011. After the initial pilot run, enhancements were made in preparation for full implementation in 2012. The MLP was reviewed every semester through the regular quality assurance mechanism of the university, and minor adjustments were made (e.g., correction of typos, reshooting the videos) throughout the years. Numerous past studies have been conducted with the MLP (see authors), thus ensuring the validity of the MLP as a learning tool and an assessment instrument.

The MLP was designed to contextualize language learning to foster an effective learning climate and allow learners to control their learning progress. The content of the activities is assessed through assignments (see the Appendix). Teachers are expected to check the progress of students through the LMS. If students are not performing well on some quizzes, teachers will offer supplementary activities to help students better grasp the content. For example, if students in some classes do not perform well in the MLP activities on academic style, teachers will always arrange more activities on academic style. As a common practice, many teachers review the performance of MLP before major assessments so that they can design some relevant revision activities before major assessments. Students are required to earn an overall score of at least 50% based on all the online quizzes.

Failure to achieve this score results in a penalty that ranges from a half- to full-grade deduction from the final grade.

3.3. Participants

This study included 17,950 students who were enrolled in a mandatory EAP course in a Hong Kong university at any point between the 2012 fall semester and the 2018 winter semester (i.e., seven academic years and 14 cohorts). The data for 2019, 2020, and 2021 were not included because class delivery was affected by social unrest and the pandemic in Hong Kong. Summer-semester students (around 30 to 60 students each year) were also not included, as the student behavioral pattern for the 7-week summer schedule was different than that observed in the other cohorts. Aside from these exceptions, this study was designed to include all students who have taken the course since its inception.

Admission to the course requires a band score of 5.48–5.56 in IELTS (International English Language Testing System) or equivalent and no prior formal training in academic literacy. Students taking this course include those in Applied Science, Business, Health, Social Sciences, and Engineering. No other demographical information was available with the data from the Learning Management System. Table 1 shows the number of students in each cohort.

Table 1. Number of students in cohorts

Semester	Number of students
2012/2013 Fall	1,730
2012/2013 Winter	397
2013/2014 Fall	1,792
2013/2014 Winter	790
2014/2015 Fall	1,695
2014/2015 Winter	620
2015/2016 Fall	2,039
2015/2016 Winter	962
2016/2017 Fall	1,867
2016/2017 Winter	884
2017/2018 Fall	1,641
2017/2018 Winter	695
2018/2019 Fall	2,246
2018/2019 Winter	592
Total	17,950

3.4. Measures

There is no simple way to operationalize self-regulated learning (Winne, 2010; Veenman et al., 2006; Rovers et al., 2019) because there is no direct measure of students' underlying mental processes. However, the adopted measures below are considered to correspond to SRL behaviors (Li et al, 2020). Still, this study considers course-based variables adopted in other studies to identify different phases of SRL (e.g., Hadwin et al., 2004; Li et al., 2020; Quic et al., 2020). In the current study, eight variables were included in the cluster analysis and further analysis to measure SRL behaviors in contextualized language learning (Table 2). It is important to note that due to the context sensitivity of SRL (Winne & Hadwin 1998), these measures were included based on how the contextualized multimodal learning was designed, and they aim to provide a generalized understanding of SRL.

The goal of this study is to identify SRL profiles and thus allow teachers and course designers to facilitate SRL based on students' SRL profiles. The results of analysis should allow teachers to take action with multimodal language learning SRL patterns. Therefore, outcome measures (i.e., course grades) were not used for cluster analysis.

Table 2. Details of measures

SRL categories	Measures	Definition	Range (before re-scaling)	Justification
Clustered measures				
Performance	Overall Course/Final Grade	Final grade in the course	0–4.5	Course Outcome
Planning	Start Day	Number of days after the start of the term that a student first submitted an MLP quiz	-2.03–96.49 (negative = starting before the term begins)	Study in Advance: suggested by Li et al. (2020)
	nth Day Before Deadline for First Attempt	Number of days before the deadline that a student first submitted an MLP quiz	-6.19–89.50 (negative = submitting a quiz after the deadline)	
	Duration	Number of days between submitting the first attempt and last attempts	0–96.45	Time between first submission and deadline adopted by Quick et al. (2020)
Performance Monitoring	Differences in Attempt Score	Difference in scores between the first and last attempt	-4.43–6.53 (negative = lower score on the last than the first attempt)	Progression of tasks suggested by Hadwin et al. (2004)
	Score in Academic Style Activities	Average score in Academic Style Activities	0–1.0	Direct outcome measure of the MLP activities
	Score in Referencing Activities	Average score in Referencing Activities	0–1.0	
	Score in Genre Activities	Average score in Genre Activities	0–1.0	
	Score in Academic Presentation Activities	Average score in Academic Presentation Activities	0–1.0	
Measures not clustered				
Performance on Assessments	Content Development	Sum of assessment scores—content development domain	0–12.5	Outcome measures correspond to MLP activities
	Organisation	Sum of assessment scores—organization domain	0–8.5	
	Language	Sum of assessment scores—language domain	0–16	
	Referencing	Sum of assessment scores—referencing domain	0–8.5	

3.5. Data processing and preparation

After retrieving the data from the learning management system, the research team processed them for cluster analysis. Students who did not complete any assessments were removed. These were not uncommon; they belonged to students who were deregistered from the course and/or the university but remained on the course list. Next, all data values were standardized (i.e., Z score) as is the common practice in cluster analysis (Sarma & Vardhan, 2018). Because cluster analysis is sensitive to outliers, the outliers were then removed. Around 2,000 data points were removed from the 14 semesters of data.

3.6. Data analysis

The objective of this study is to identify profiles and patterns of learners' SRL with no pre-existing assumptions or expected profile. Therefore, cluster analysis was adopted. Cluster analysis is an exploratory technique and should not be treated as an "outcome practice" for hypothesis testing (Sarma & Vardhan, 2018). Using the final data set, the number of clusters was determined using the "elbow method," identifying the dipping/changing point from the Total Sum of Within Squares (Bholowalia & Kumar, 2014). After that, the k-means cluster analysis was conducted with R (version 4.0.3). Then, the overall average of the silhouette values was examined as an indicator of cohesion and separation (Hao et al., 2021). This measure can range from -1 to 1. A positive measure is desirable.

After completing the cluster analysis, a MANOVA (with IBM SPSS Statistics 27) was used to explore the differences between clusters, using the cluster groups as the grouping variables and the SRL behaviors and performance variables as dependent variables. It is important to note that MANOVA was used only to explore the extent of the between-cluster differences across the indicators, as the indicators were expected to be different after being clustered. This aligns with the methods used in another SRL study (Ng et al., 2016; Broadbent & Fuller-Tyszkiewicz, 2018). The alpha value was set at 0.05.

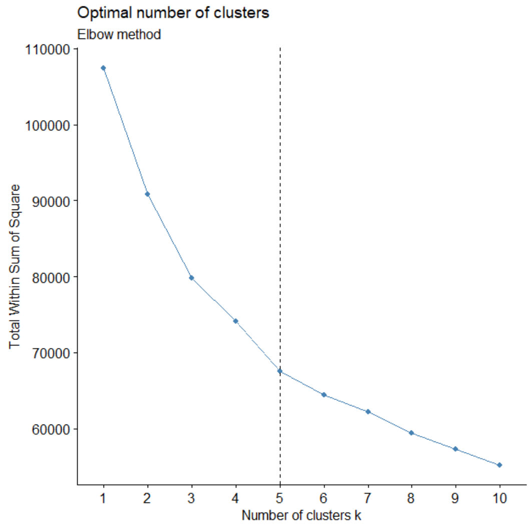
4. Results

The objective of this study was to explore students' SRL behaviors. Cluster analysis was first adopted with the elbow method to find the optimal number of clusters, followed by the k-means clustering technique and then MANOVA to determine if indicators of SRL differed among the clusters. The characteristics of the clusters (i.e., profiles) were then described and discussed.

4.1. Step 1 – Optimal number of clusters and k-means clustering

To determine the optimal number of clusters, the elbow method was adopted. This method is based on the Total Sum of Within Squares and can be represented graphically. Figure 1 shows that four clusters are the optimal number.

Figure 1. Optimal number of clusters with elbow method



Five clusters were subsequently formulated (see Table 3 for the descriptive statistics). While there is no consensus on a fit index for cluster analysis, and most measures were used for comparisons, silhouette values were used to assess indicate the adequacy of the cluster analysis. The average of the silhouette measures was 0.14, which suggested that clustering was still desirable (Hao et al., 2021). Appendix 2 presents a Radar chart as a visualizations of all clusters.

Table 3. Descriptive statistics for clusters

	Cluster 1 (n = 3974)		Cluster 2 (n = 2127)		Cluster 3 (n = 2324)		Cluster 4 (n = 4876)		Cluster 5 (n = 2713)	
Clustered measures										
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Overall course grade	0.00	0.84	-0.43	0.82	0.30	0.85	0.30	0.84	0.09	0.80
Start day	-0.60	0.56	0.78	0.89	-0.62	0.45	-0.42	0.68	1.22	0.47
Nth Day before Deadline for First Attempt	-0.36	0.59	-0.48	0.58	1.62	0.65	0.15	0.77	-0.66	0.48
Duration	0.38	0.62	0.02	0.93	-0.72	0.90	0.73	0.64	-0.95	0.75
Differences in Attempt Score	-0.10	0.51	0.12	0.81	0.00	0.78	-0.08	0.64	-0.35	0.54
Score in Academic Style Activities	0.28	0.56	-0.41	0.89	0.14	0.77	0.34	0.64	0.22	0.70
Score in Referencing Activities	0.26	0.55	-0.29	0.81	0.24	0.56	0.30	0.60	0.27	0.52
Score in Genre Activities	-0.07	0.85	-1.02	0.95	0.09	0.73	0.58	0.59	0.29	0.60
Score in Academic Presentation Activities	-0.89	0.44	0.44	0.70	-0.52	0.96	0.93	0.34	-0.31	0.98
Measures not clustered										
Content development	-0.04	0.91	-0.35	0.90	0.25	0.88	0.23	0.91	0.06	0.89
Organisation	-0.04	0.92	-0.39	0.90	0.23	0.92	0.24	0.94	0.06	0.91
Language	-0.02	0.90	-0.31	0.91	0.14	0.99	0.16	0.95	0.09	0.91
Referencing	0.02	0.90	-0.42	0.90	0.29	0.90	0.25	0.91	-0.02	0.93

4.2. Step 2 – Exploration of SRL profiles

MANOVA was conducted to explore further the differences and similarities across the four clusters. Dependent variables included all the SRL indicators used in clustering, along with the content development, organization, language, referencing, and presentation assessment outcomes. MANOVA was in no way a validation of the clusters, as clustering is an exploratory technique for identifying patterns, not an outcome process for testing hypotheses (Sarma & Vardhan, 2018). However, MANOVA is useful for exploring the differences across clusters. DiFrancesca et al. (2016) used MANOVA similarly as a follow-up technique in their SRL study.

There was a statistically significant difference in all SRL indicators and assessments outcomes based on clusters, $F(52, 61958.23) = 1565.77$, $p < .05$; Wilk's $\Lambda = 0.039$, partial $\eta^2 = .56$. Further univariate ANOVA indicated significant main effect of clusters on all indicators and outcomes. Table 4 shows the univariate ANOVA results. For the descriptive statistics of the clusters, see Table 3.

The purpose of this study was to identify how EAP learners use SRL in regard to multimodal formative quizzes in a large EAP course at a higher education institution in Hong Kong. Our data, based on descriptive statistics, revealed two major findings. First, the students fell into five clusters: two groups that performed well while exhibiting different SRL behaviors; two groups that performed at par while exhibiting different SRL behaviors; and one group that performed poorly and exhibited few SRL behaviors. Second, we found that the identification of five clusters of students and their behaviors in completing the multimodal formative quizzes confirmed the importance of SRL (e.g., goal setting, time-management) (see Dörnyei & Ryan, 2015) in language learning. See Table 5 for a summary of the results.

Table 4. Follow-up ANOVA for clustered and un-clustered variables

	df	*F* value	Partial eta squared
Clustered measures			
Overall course grade	4	229.03*	0.08
Start day	4	2107.03*	0.58
Nth day before deadline for first attempt	4	2112.06*	0.56
Duration	4	1678.31*	0.43
Differences in attempts	4	72.88*	0.04
Scores in academic style activities	4	225.62*	0.11
Scores in referencing activities	4	149.57*	0.09
Score in genre activities	4	991.05*	0.31
Score in academic presentation activities	4	2152.31*	0.55
Measures not clustered			
Content development	4	191.54*	0.05
Organization	4	199.48*	0.05
Language	4	108.58*	0.03
Referencing	4	243.25*	0.06

Note. $^{*}p < .05$.

Table 5. Summary of cluster characteristics

Categories (indicators)	Cluster 1 (*n* = 3974)	Cluster 2 (*n* = 2127)	Cluster 3 (*n* = 2324)	Cluster 4 (*n* = 4876)	Cluster 5 (*n* = 2713)
Course performance	At par	Poor performance	Good performance	Good performance	At par
Planning					
Start day	Started early	Started late	Started early	Started later / not the earliest	Started late
Nth day before deadline for first attempt	Started shortly before the deadline	Started shortly before the deadline	Started shortly before the deadline		Started shortly before the deadline
Performance monitoring					
Scores on activities	Above mean on most activities, except the last one	Below mean on most activities, except the last one	Above mean scores on most activities, except the last	Above mean on most activities	Above mean on most activities, except the last one
Differences between attempts		Improved between attempts			Did not improve much between attempts
Duration	Spent more time	At par	Spent less time	Spent more time	Spent less time

5. Discussion

5.1. Five types of SRL behaviors

We identified five clusters of students who differed distinctly in their SRL behaviors in completing the multimodal formative quizzes. The EAP students in Cluster 3 and Cluster 4 earned the highest grades on the quizzes, and those in Cluster 2 received the lowest scores. Students in Cluster 4 and Cluster 5 performed at par (i.e., close to 0 for standardized score).

As the results for the students in Cluster 3 and Cluster 4 demonstrate, there was a correlation between success aided by starting the quizzes well before the deadline and taking the multimodal quizzes seriously (i.e., obtaining above-average scores on most quizzes). In these two clusters, students may start later or early and may spend less time or more time on quizzes. The grades of the students in Cluster 1 and Cluster 5 were slightly lower than those in Cluster 3 and Cluster 4. Students in these clusters began the quizzes shortly before the deadline but demonstrated positive SRL behaviors by taking the quizzes seriously (i.e., obtaining above-average marks in their first attempt at most quizzes). The main difference between these two clusters (i.e., good performance and at-par performance) lay in the matter of how long before the deadline the students began their quizzes. Students who performed well began their multimodal quizzes well before the deadline, but students who performed at par began only shortly before the deadline. It is important to note that students in these four clusters, achieving at least at-par performance if not better, took the multimodal quizzes seriously. This finding supports the argument that students who display self-regulatory behaviors are more prone to monitoring and adjusting their learning processes (Cho & Cho, 2017). It also accords with Azevedo and Hadwin's (2005) observation that there is a strong correlation between SRL and academic success. This study, however, finds that the SRL behavior that matters is how long before a deadline students begin and whether they monitor their learning.

The students in Cluster 2 demonstrated poor skills in SRL, beginning their activities late and close to the deadlines. Although their activity scores were below the mean for almost all quizzes, they did not monitor their progress or improve their scores. These students did not perform well on course assessments. Thus, this group of students did not take advantage of the multimodal formative online quizzes to obtain ongoing and timely feedback on their learning (see more discussion from Nicole and Macfarlane-Dick, 2006 on how students can take more responsibility in their learning). By completing formative quizzes, EAP students can discover their weaknesses and take steps to mitigate them before summative assessments take place. Formative quizzes are also important tools for teachers to monitor students' understanding of language concepts (Gardner, 2012; Hinkelman, 2018). As the results indicated, Cluster 2 students did not use the quizzes to revise and consolidate knowledge. Nor did the quizzes develop SRL, contrary to McLaughlin and Yan's (2017) who found that such quizzes can improve SRL of students. Because of their lack of SRL behaviors and poor performance, the individuals in Cluster 2 could be classified as unsuccessful students (Gerami & Baighlou, 2011).

The students in Cluster 5 demonstrated a pragmatic form of SRL by doing just enough. This group began the quizzes later and performed well in the first few activities. Then, their performance declined, and they did not improve much between attempts at the quizzes. However, their course assessments were right at par. This may suggest that they did not demonstrate enough SRL to do better, but it could also signify that their SRL skills were outstanding. Students in this group could have planned well, putting in enough effort on the earlier activities to stop doing activities later. They may have allocated only enough time and effort to complete the quizzes, thus displaying outstanding SRL. Although the Cluster 5 students performed at par in the course, they demonstrated as few SRL behaviors as the Cluster 2 students. They were reluctant to make an effort to learn and could be classified as passive students (Zimmerman & Schunk, 2011).

While good mastery of SRL was strongly related to course outcomes in the contextualized language course, we observed only more obvious links between SRL and some assessment components (i.e., content development and referencing), but not other assessment components (i.e., organization and referencing). One possible explanation is that SRL reflects the overall effort put into the course (Cho & Heron, 2015), and this can only be revealed by certain assessment components. For example, content development assessment focuses on the research that students have conducted for their essays (e.g., searching for sources, reading, and developing strong arguments). The more effort students expend on content development, the better the grade they will achieve, which should, to a great extent, reflect the effort students put in for MLP activities. The same applies to conventions that require students to prepare in-text citations and the reference list carefully; these require students to make efforts to check style guides to ensure that their citations are correct (e.g., formatting and punctuation).

In contrast, some assessment components (i.e., organization and language) may not accurately reflect the content and effort put into MLP activities. For example, academic style is included in the MLP activities and in language assessments, but the effort put into the use of varied and accurate language, which is central to language assessment, is not reflected in MLP activities. In the same vein, the assessment of organization is related to subtle genre knowledge and rhetorical skills. While there are MLP activities addressing genre knowledge, such as organizational structure of essays, students can easily acquire such genre knowledge by referring to essay samples or class notes regardless of their performance in MLP activities. Students may perform well because they can write logically when they complete the assessment; therefore, effort and content in MLP genre and activities may not be strong indicators. More investigations into the correlation between MLP SRL, contextualized language learning, and overall assessment performance should be performed.

6. Conclusion

This study provides insights into the relationship between students' SRL and performance through an analysis of data from multimodal formative quizzes. The results revealed correlations between SRL and academic performance, planning, and performance monitoring. It showed that students who start well before deadlines and take formative quizzes seriously will perform best. This study also shows five different profiles of students when they complete formative online quizzes. Based on our findings, we recommend that educators develop personalized instructions for each cluster of students to motivate, stimulate, and foster SRL. This should improve course performance. Having access to learner analytics allows educators to adjust blended, multimodal formative quizzes to meet the needs and interests of specific student cohorts (Ferguson, 2012). As educators, we need to understand and address EAP students' needs to improve the flexibility and efficiency of their blended learning experiences.

6.1. Limitations and future studies

Several limitations of this study should be noted. Although this study provides evidence of a relationship between SRL and strong academic performance, all of the respondents came from one university, thus limiting the generalizability of our findings. In addition, clustering is an exploratory technique; it is not designed to assess outcomes (e.g., hypothesis testing). Also, model adequacy measures indicated that the cluster structure was weak. These limit the validity of claims made about SRL behaviors.

We propose several topics for future study. Researchers could focus on gathering rich information by using both a questionnaire and interviews to complement the objective quiz data. Such qualitative data might provide insight into students' perceptions of the interfaces and designs of the multimodal formative quizzes and their levels and sources of motivation. It could also suggest ways that the quality of the quizzes could be enhanced using the features of LMS platforms. Finally, research in educational settings beyond Asia could shed more light on the relationships among SRL, multimodal formative quizzes, and strong course performance. We hope that the findings of this study serve to remind educators that EAP students need to develop strong SRL skills and engage seriously in multimodal contextualized language learning to succeed in their studies.

References

Antonenko, P. D., Toy, S., & Niederhauser, D. S. (2012). Using cluster analysis for data mining in educational technology research. *Educational Technology Research and Development*, *60*(3), 383-398. https://link.springer.com/article/10.1007/s11423-012-9235-8

Artino, A. R. (2007). Motivational beliefs and perceptions of instructional quality: Predicting satisfaction with online training. *Journal of Computer Assisted Learning*, *24*(3), 260-270. https://doi.org/10.1111/j.1365-2729.2007.00258.x

Azevedo, R., & Hadwin, A. F. (2005). Scaffolding self-regulated learning and metacognition – implications for the design of computer-based scaffolds. *Instructional Science, 33*, 367-379. https://doi.org/10.1007/s11251-005-1272-9

Bholowalia, P., & Kumar, A. (2014). EBK-means: A Clustering technique based on elbow method and k-means in WSN. *International Journal of Computer Applications*, *105* (9), 17–24.

Boekaerts, M. (1999). Motivated learning: Studying student situation transactional units. *European Journal of Psychology of Education*, *14*(1), 41-555. https://doi.org/10.1007/BF03173110

Bol, L., & Garner, J.K. (2001). Challenges in supporting self-regulation in distance education environment. *Journal of Computing in Higher Education, 23*(2–3) (2011),104-123. https://doi.org/10.1007/s12528-011-9046-7

Broadbent, J. (2017). Comparing online and blended learner's self-regulated learning strategies and academic performance. *The Internet and Higher Education, 33*, 24-32. https://doi.org/10.1016/j.iheduc.2017.01.004

Broadbent, J., & Poon, W. L. (2015). Self-regulated learning strategies & academic achievement in online higher education learning environments: A Systematic review. *The Internet and Higher Education, 27,* 1-13. https://doi.org/10.1016/j.iheduc.2015.04.007

Broadbent, J., & Fuller-Tyszkiewicz, M. (2018). Profiles in self-regulated learning and their correlates for online and blended learning students. *Educational Technology Research and Development, 66*(6), 1435-1455. https://link.springer.com/article/10.1007/s11423-018-9595-9

Chapelle, C. A., & Sauro, S. (2017). *The Handbook of Technology and Second Language Teaching and learning*. Wiley-Blackwell. https://doi.org/10.1002/9781118914069

Cho, M. H., & Heron, M. L. (2015). Self-regulated learning: The Role of motivation, emotion, and use of learning strategies in students' learning experiences in a self-paced online mathematics course. *Distance Education, 36*(1), 80–99. https://doi.org/10.1080/01587919.2015.1019963

Cho, M. H., & Jonassen, D. (2009). Development of the human interaction dimension of the self-regulated learning questionnaire in asynchronous online learning environments. *Educational Psychology, 29*(1), 117-138. https://doi.org/10.1080/01443410802516934

Cho, M. H., & Shen, D. (2013). Self-regulation in online learning. *Distance Education, 34*(3), 290-301. https://doi.org/10.1080/01587919.2013.835770

Cho, M. H., & Cho, Y. (2017). Self-regulation in three types of online interaction: a scale development. *Distance Education, 38*(1), 70-83. https://doi.org/10.1080/01587919.2017.1299563

Cole, A. W., Lennon, L., & Weber, N. (2021). Student perceptions of online active learning practices and online learning climate predict online engagement. Interactive *Learning Environments, 29*(5), 866-880. https://doi.org/10.1080/10494820.2019.1619593

Cook, B. R., & Babon, A. (2017). Active learning through online quizzes: better learning and less (busy) work. *Journal of Geography in Higher Education, 41*(1), 24-38. https://doi.org/10.1080/03098265.2016.1185772

Coskuncay, F., & Ozkan, S. (2013). A Model for instructors' adoption of learning management systems: Empirical validation in higher education context. *Turkish Journal of Educational Technology, 12*(2), 13-25. http://tojet.net/articles/v12i2/1222.pdf

DiFrancesca, D., Nietfeld, J. L., & Cao, L. (2016). A Comparison of high and low achieving students on self-regulated learning variables. *Learning and Individual Differences, 45*, 228-236. https://doi.org/10.1016/j.lindif.2015.11.010

Dörnyei, Z., & Ryan, S. (2015). *The Psychology of the language learner revisited.* Routledge.

Dörnyei, Z. (2005). *The Psychology of the language learner: Individual differences in second language acquisition*. Lawrence Erlbaum.

ElSayed, A. A., Caeiro-Rodríguez, M., MikicFonte, F. A., & Llamas-Nistal, M. (2019). Research in learning analytics and educational data mining to measure self-regulated learning: A Systematic review. In *World conference on mobile and contextual learning* (pp. 46-53). https://www.learntechlib.org/p/210600/

Ferguson, R. (2012). Learning analytics: Drivers, developments and challenges. *International Journal of Technology Enhanced Learning, 4*(5-6), 304–317. https://doi.org/10.1504/IJTEL.2012.051816

Gamage, S. H. P. W., Ayres, J. R., Behrend, M. B., & Smith, E. J. (2019). Optimising Moodle quizzes for online assessments. *International Journal of Stem Education, 6*(1). https://doi.org/10.1186/s40594-019-0181-4

Gardner, J. (2012). *Assessment and learning*. Sage. https://dx.doi.org/10.4135/9781446250808

Garrison, D. R., & Vaughan, N. D. (2013). Institutional change and leadership associated with blended learning innovation: Two case studies. *Internet and Higher Education, 18*(1), 24-28. https://doi.org/10.1016/j.iheduc.2012.09.001

Gerami, M. H., & Baighlou, S. M. G. (2011). Language learning strategies used by successful and unsuccessful Iranian EFL students. *Procedia – Social and Behavioral Sciences, 29*, 1567-1576. https://doi.org/10.1016/j.sbspro.2011.11.399

Guo, S., Wang, X., Deng, W., Hong, J., Wang, J., & Wu, Y. (2022). Whose spatial ability benefits from learning with 3D design? From the perspective of learning analysis. *Educational Technology & Society, 25*(1), 179-192.

Hafner, C. A., & Miller, L. (2011). Fostering learner autonomy in English for science: A Collaborative digital video project in technological learning environment. *Language Learning and Technology, 15*, 68–86. https://doi.org/10125/44263

Hafner, C. A., & Pun, J. (2020). Editorial: Introduction to this special issue: English for Academic and Professional purposes in the digital era. *RELC Journal, 51*(1), 3-13. https://doi.org/10.1177/0033688220917472

Hampel, R., & Hauck, M. (2006). Computer-mediated language learning: Making meaning in multimodal virtual learning spaces. *The JALT CALL Journal, 2*(2), 3-18.

Hadwin, A., Boutara, L., Knoetzke, T., & Thompson, S. (2004). Cross-case study of self-regulated learning as a series of events. Educational Research and Evaluation, 10(4-6), 365-417. https://doi.org/10.1080/13803610512331383499

Hao, R., Zhang, L., Liu, J., Liu, Y., Yi, J., & Liu, X. (2021). A Promising approach: Artificial intelligence applied to Small Intestinal Bacterial Overgrowth (SIBO) diagnosis using cluster analysis. *Diagnostics, 11*, Article No. 1445. https://doi.org/10.3390/diagnostics11081445

Hill, J., & Hannafin, M. (2001). Teaching and learning in digital environments: The Resurgence of resource-based learning. *Educational Technology, Research and Development, 49*, 37-52.

Hinkelman, D. (2018). *Blended technologies in second language classrooms.* Palgrave Macmillan. https://link.springer.com/book/10.1057/978-1-137-53686-0

Jiang, L., & Ren, W. (2021). Digital multimodal composing in L2 learning: Ideologies and impact. *Journal of Language, Identity & Education, 20*(3), 167-182. https://doi.org/10.1080/15348458.2020.1753192

Kendrick, M., Jones, S., Mutonyi, H., & Norton, B. (2006). Multimodality and English education in Ugandan schools. *English Studies in Africa, 49*(1), 95-114.

Kohnke, L., Foung, D., & Chen, J. (2022). Using learning analytics to explore the potential contribution of multimodal formative assessments to academic success in higher education. *SAGE Open*, *12* (2), 1-12. https://doi.org/10.1177/21582440221089957

Kohnke, L., & Jarvis, A. (2022). Developing infographics for English for academic purposes courses. *TESOL Journal*. https://doi.org/10.1002/tesj.675

Kohnke, L., Jarvis, A., & Ting, A. (2021). Digital multimodal composing as authentic assessment in discipline-specific English courses: Insights from ESP learners. *TESOL Journal*, *12*(3), 1-15. https://doi.org/10.1002/tesj.600

Kress, G., & van Leeuwen, T. (2001). *Multimodal discourse: The Modes and media of contemporary communication*. Edward Arnold.

Kuo, Y. C., Walker, A., Belland, B. & Schroder, K. (2013). A Predictive study of student satisfaction in online education programs. *The International Review of Research in Open and Distributed Learning, 14*(1), 16-39. https://doi.org/10.19173/irrodl.v14i1.1338

Lamy, M. (2018). Diversity in modalities. In G. Stockwell (Ed.), *Computer-assisted language learning: Diversity in research and practice* (pp. 109–126). Cambridge University Press.

Lawanto, O., Goodridge, W. H., Santoso, H., & Lawanto, K. (2014). Task-value, self-regulated learning, and performance in a web-intensive undergraduate engineering course: How are they related? *Journal of Online Learning and Teaching, 10*(1), 97-111.

Lehmann, T., Hähnlein, I., & Ifenthaler, D. (2014). Cognitive, metacognitive and motivational perspectives on preflection in self-regulated online learning. *Computers in Human Behavior, 32*, 313–323. https://doi.org/10.1016/j.chb.2013.07.051

Li, Q., Baker, R., & Warschauer, M. (2020). Using clickstream data to measure, understand, and support self-regulated learning in online courses. *The Internet and Higher Education, 45*, Article 100727. https://doi.org/10.1016/j.iheduc.2020.100727

Marcel, F. (2020). *Mobile mixed reality technologies for language teaching and learning* (Unpublished doctoral dissertation). University of Toronto, Canada. https://hdl.handle.net/1807/103360

Martin, D. (2020). Providing Students with multimodal feedback experiences. *Journal of Curriculum, Teaching, Learning and Leadership in Education, 5*(1), 16-27.

Mauricio, M. C., & Genuino, C. F. (2020). Writing performance and attitude of ESL learners engaged in smartphone assisted collaborative activity. *LEARN Journal: Language Education and Acquisition Research Network, 13*(2), 445-468.

McLaughlin, T., & Yan, Z. (2017). Diverse delivery methods and strong psychological benefits: A Review of online formative assessments. *Journal of Computer Assisted Learning, 33*, 562-574. https://doi.org/10.1111/jcal.12200

Ng, B. L., Liu, W. C., & Wang, J. C. (2016). Student motivation and learning in mathematics and science: A Cluster analysis. *International Journal of Science and Mathematics Education, 14*(7), 1359-1376. https://doi.org/10.1007/s10763-015-9654-1

Nicole, D., & Macfarlane-Dick, D. (2006). Formative assessment and self-regulated learning: A Model and seven principles of good feedback practice. *Studies in Higher Education, 31*, 199–218. https://doi.org/10.1080/03075070600572090

Owston, R., & York, D. N. (2018). The Nagging question when designing blended courses: Does the proportion of time devoted to online activities matter. *The Internet and Higher Education, 36*, 22-32. https://doi.org/10.1016/j.iheduc.2017.09.001

Pachler, N., Daly, C., Mor, Y., & Mellar, H. (2010). Formative e-assessment: Practitioner cases. *Computers & Education, 54*, 715–721. https://doi.org/10.1016/j.compedu.2009.09.032

Padilla-Walker, L. M. (2006). The Impact of daily extra credit quizzes on exam performance. *Teaching of Psychology, 33*(4), 236-239. https://doi.org/10.1207/s15328023top3304_4

Paivio, A. (1986). *Mental representations: A Dual coding approach.* Oxford University Press. https://academic.oup.com/book/10932

Park, Y., Yu, J. H., & Jo, I.-H. (2016). Clustering blended learning courses by online behavior data: A Case study in a Korean higher education institute. *The Internet and Higher Education, 29*, 1-11. https://doi.org/10.1016/j.iheduc.2015.11.001

Peters, E. (2019). The Effect of imagery and on-screen text on foreign language vocabulary learning from audio-visual input. *TESOL Quarterly, 53*(4), 1008–1032. https://doi.org/10.1002/tesq.531

Pintrich, P.R. (1995). Understanding self-regulated learning. *New Directions for Teaching & Learning, 63*, 3-12.

Pintrich, P. R. (2004). A Conceptual framework for assessing motivation and self-regulated learning in college students. *Educational Psychology Review, 16*, 385-407.

Porter, W. W., Graham, C. R., Bodily, R. G., & Sandberg, D. S. (2016). A Qualitative analysis of institutional drivers and barriers to blended learning adoption in higher education. *The Internet and Higher Education, 28,* 17-27.

Pujadas, G., & Muñoz, C. (2019). Extensive viewing of captioned and subtitled TV series: A Study of L2 vocabulary learning by adolescents. *The Language Learning Journal, 47*(4), 479–496. https://doi.org/10.1080/09571736.2019.1616806

Quick, J., Motz, B., Israel, J., & Kaetzel, J. (2020). What college students say, and what they do: Aligning self-regulated learning theory with behavioral logs. In *Proceedings of the Tenth International Conference on Learning Analytics & Knowledge* (pp. 534-543). https://doi.org/10.1145/3375462.3375516

Ross, B., Chase, A. M., Robbie, D., Oates, G., & Absalom, Y. (2018). Adaptive quizzes to increase motivation, engagement and learning outcomes in a first year accounting unit. *International Journal of Educational Technology in Higher Education, 15*(30). https://doi.org/10.1186/s41239-018-0113-2

Ross, B., & Gage, K. (2006). Global perspectives on blended learning: Insights from WebCT and our customers in higher education. In C. J. Bonk, & C. R. Graham (Eds.), *Handbook of blended learning: Global perspectives, local designs* (pp. 155-168). Pfeiffer.

Rovers, S. F., Clarebout, G., Savelberg, H. H., de Bruin, A. B., & van Merriënboer, J. J. (2019). Granularity matters: Comparing different ways of measuring self-regulated learning. *Metacognition and Learning, 14*(1), 1-19. https://link.springer.com/article/10.1007/s11409-019-09188-6

Sarma, K. V. S., & Vardhan, R. V. (2018). Cluster analysis and its applications. In K. V. S. Sarma, & R. V. Vardhan (Eds.), *Multivariate Statistics Made Simple: A Practical Approach* (pp. 219-237). Taylor and Francis. https://doi.org/10.1201/9780429465185

Shin, T., & Cimasko, D. S. (2008). Multimodal composition in a college ESL Class: New tools, traditional norms. *Computers & Composition, 25*(4), 376-395. https://doi.org/10.1016/j.compcom.2008.07.001

Schunk, D. H., & Pajares, F. (2002). The Development of academic self-efficacy. In A. Wigfield & J. S. Eccles (Eds.), *Development of achievement motivation* (pp. 15–31). Academic Press. https://doi.org/10.1016/B978-012750053-9/50003-6

Stenlund, T., Lyrén, P. E., & Eklöf, H. (2018). The Successful test taker: Exploring test-taking behavior profiles through cluster analysis. *European Journal of Psychology of Education, 33*(2), 403-417. https://link.springer.com/article/10.1007/s10212-017-0332-2

Tragant, E., & Pellicer-Sánchez, A. (2019). Young EFL learners' processing of multimodal input: Examining learners' eye movements. *System, 80*, 212–223. https://doi.org/10.1016/j.system.2018.12.002

Tseng, W.-T., Dörnyei, Z., & Schmitt, N. (2006). A New approach to assessing strategic learning: The Case of self-regulation in vocabulary acquisition. *Applied Linguistics, 27*(1), 78–102. https://doi.org/10.1093/applin/ami046

Veenman, M. V. J., Van Hout-Wolters, B. H. A. M., & Afflerbach, P. (2006). Metacognition and learning: Conceptual and methodological considerations. *Metacognition and Learning, 1*(1), 3–14. https://doi.org/10.1007/s11409-006-6893-0

Wandler, J., & Imbriale, W. (2017). Promoting undergraduate student self-regulation in online learning environments. *Online Learning 21*(2). https://doi.org/10.24059/olj.v21i2.881

Wang, C., Schwab, G., Fenn, P., & Chang, M. (2013). Self-efficacy and self-regulated learning strategies for English language learners: Comparison between Chinese and German college students. *Journal of Educational and Developmental Psychology, 3*(1), 173–191. https://doi.org/10.5539/jedp.v3n1p173

White, C., Zheng, Y., & Skyrme, G. (2021). Developing a model for investigating one-to-one synchronous Chinese online language teaching via videoconferencing. *Computer Assisted Language Learning, 34*(1-2), 92-113. https://doi.org/10.1080/09588221.2020.1770800

Wilkie, B., & Liefeith, A. (2020). Student experiences of live synchronised video feedback in formative assessment. *Teaching in Higher Education*, *27*(3), 403-416. https://doi.org/10.1080/13562517.2020.1725879

Winne, P. H. (2010). Improving measurement of self-regulated learning. *Educational Psychologist, 45*(4), 267-276. https://doi.org/10.1080/00461520.2010.517150

Winne, P. H., & Hadwin, A. F. (1998). Studying as self-regulated learning. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *Metacognition in educational theory and practice* (pp. 277–304). Lawrence Erlbaum.

Winne, P. H., & Perry, N. E. (2000). Measuring self-regulated learning. In M. Boekaerts, P. R. Pintrich, & M. Zeidner (Eds.), *Handbook of self-regulation* (pp. 531-566). Academic Press. https://doi.org/10.1016/B978-012109890-2/50045-7

Yeh, H.C. (2018). Exploring the perceived benefits of the process of multimodal video making in developing multiliteracies. *Language Learning & Technology, 22*(2), 28–37. https://doi.org/10125/44642

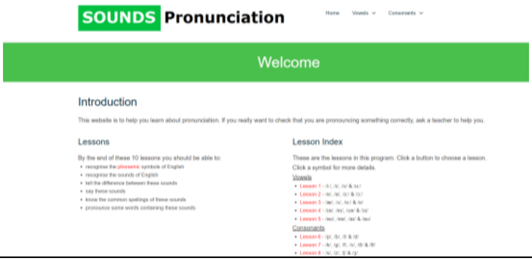
Yuan, C., & Yang, H. (2019). Research on K-value selection method of K-means clustering algorithm. *J Multidisciplinary Scientific Journal, 2*(2), 226-235. https://doi.org/10.3390/j2020016

Zimmerman, B. J. (1990). Self-regulated learning and academic achievement: an overview. *Educational Psychologist, 25*(1), 3-17. https://doi.org/10.1207/s15326985ep2501_2

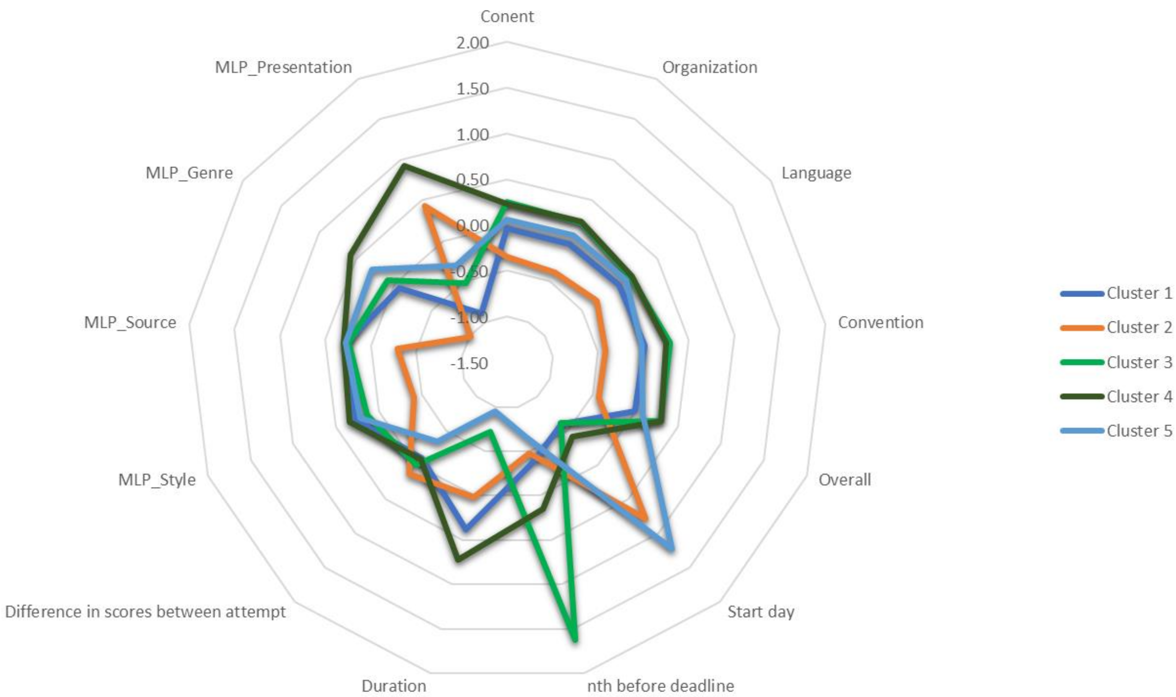
Zimmerman, B. J., & Schunk, D. H. (2011). Self-regulated learning and performance: An Introduction and overview. In B. J. Zimmerman, & D. H. Schunk (Eds.), *Handbook of self-regulation of learning and performance* (pp. 1–15). Routledge.

Appendix 1 – Examples of Multimodal Learning Package (MLP)

List of 13 Quizzes	Categories	Examples of Multimodal Content	Examples of Formative Quiz	Corresponding Assessment Criteria
Session 1 “Academic Style”	Academic Style	Website [URL for institution removed]: Screenshot to be displayed for peer review; link to be displayed in the final publication  Infographics for Academic style	Drag and Drop activity, on definitions of academic style issues (e.g., contraction)	Language (All writing assessments)
Session 2 “Referencing” Session 4 “Integrating Sources”	Referencing	Youtube [name of institution removed]: Screenshot to be displayed for peer review; link to be displayed in the final publication 	MC questions on content presented in the videos	Referencing (All writing assessments)
Session 3 “Reading Academic Articles” Session 5 “Essay Writing (1)” Session 6 “Essay Writing (2)” Session 7 “Fact vs Opinion” Session 8 “‘For’ &	Genre Knowledge	Youtube [name of institution removed] 	MC questions, on facts presented in videos, problems with Introduction paragraph, flow of an Introduction paragraph	Content Development / Organization (All writing assessments)

‘Against’ Essays” Session 9 “Editing your Work”				
Session 10 “Academic Presentations” Session 11 “Visual Aids” Session 12 “Effective Presentation Delivery” Session 13 “Presentations ”	Presentation Skills	Youtube [name of institution removed]  Pronunciation Website hosted on by the institution: Screenshot to be displayed for peer review; link to be displayed in the final publication 	Labelling activity, on how to create interest	All criteria for Presentation assessment

Appendix 2 – Visualization of Five Clusters



Muntaha, M., Chen, J., & Dobinson, T. (2023). Exploring Students' Experiences of Using Multimodal CMC Tasks for English Communication: A Case with Instagram. *Educational Technology & Society*, *26*(3), 69-83. https://doi.org/10.30191/ETS.202307_26(3).0006

Exploring Students' Experiences of Using Multimodal CMC Tasks for English Communication: A Case with Instagram

Muntaha Muntaha*, Julian Chen and Toni Dobinson
School of Education, Curtin University, Perth, Australia // muntaha.abdulqodir@postgrad.curtin.edu.au // julian.chen@curtin.edu.au // T.Dobinson@Curtin.edu.au
*Corresponding author

ABSTRACT: Employing multimodal computer-mediated communication (CMC) for online language learning and teaching has gained momentum worldwide due to the emergence of various digital modes, such as text, image, audio, and video, for online communication. This pilot study aimed to explore students' learning experiences with multimodal CMC tasks through Instagram. Thirty first-year students at an Indonesian university completed seven CMC tasks, consisting of information gap, reasoning gap, and opinion gap tasks, through three Instagram communication channels: text chat, voice chat, and video chat. Pre- and post-study surveys, journal reflections, and interviews were analyzed using a mixed methods approach. Findings revealed that students overall positively perceived their experiences with tasks delivered through Instagram video, audio, and text chats. They also reported that paralinguistic features afforded by the multimodal Instagram channels—such as emojis, GIFs, images in text chat, intonation in voice chat, and gestures in video chat—facilitated effective communication. However, challenges such as poor internet connections, lack of consciousness of student agency when interacting in video chats, and high anxiety at the beginning of task implementation were also documented during student task performance. The findings suggest that the use of multimodal CMC channels affords greater accessibility and provides multimodal affordances for language learners to communicate using rich semiotic resources. They can strategically draw upon their digital literacy skills to convey messages during meaningful task interaction. Nevertheless, language instructors should consider the availability of internet infrastructure and students' language proficiency prior to utilizing multimodal CMC channels as language learning tools.

Keywords: Multimodality, CMC, Social networking sites, TBLT, Instagram

1. Introduction

The emergence of digital technology has significantly equipped CMC with multimodal features in which users may utilize various modes to convey messages. In the current context of language learning, multimodal CMC has become an indispensable part of learning, evidenced in a variety of widely used communication platforms, such as email, social networking sites, and videoconferencing (Wigham & Satar, 2021). Due to the nature of multimodal CMC, language learners are able to tap into various modes to help convey meaning by mixing text, audio, images, and emoticons. This is achieved, for example, by the use of text chat during videoconferencing, which helps learners to contribute to the discussion board without interrupting the speaker (Hampel & Stickler, 2012). Images and/or emoticons can assist lower-level English language learners to be able to actively engage in text chat conversations (Jin, 2017). The interplay of multimodal communication further enriches learning experiences and facilitates language production beyond a single mode of communication. As a result, the use of multimodal environments in online language learning, and affordances in different types of CMC, have been identified as key research areas in computer-assisted language learning (CALL) research and language learning practices (Guichon & Cohen, 2016).

Social networking sites (SNSs), as the latest revolution of CMC, have become one of the most common platforms used for assisting language learning and teaching (Reinhardt, 2018). To date, Instagram, an application that supports image, text, audio, and video sharing, is one of the most popular SNSs around the world besides Facebook, Twitter, and TikTok (Walsh, 2021). It enables users to send text and audio messages and hold video calls with a single user or group. In the Indonesian context, SNSs are used for socializing with a wider community and for everyday communication and interaction. For example, as an evolving SNS platform, Instagram was reported to have more than 91 million users in Indonesia in 2021 (NapoleonCat, 2021). Notably, around 80% of Indonesian users were teenagers and young adults between 13 and 34 years old (Nurhayati-Wolff, 2021b) actively and frequently using Instagram for daily social life. Indonesian young adults position Instagram as a means for completing social tasks, such as making a meeting appointment, sharing and discussing personal feelings, or promoting commercial sales for personal services and products (Prihantoro & Zulizilah, 2017).

Currently, the majority of literature on the use of Instagram for language learning has only focused on: (1) exploring Instagram features for posting/uploading learning materials, (2) asking the students to respond to the posting by discussing content shared, (3) providing examples in the comments section (e.g., Park & Wu, 2020; Yudhiantara & Nuryantini, 2019). Little attention has been paid to the affordance of synchronous CMC (SCMC) and multimodal features of Instagram, especially for completing authentic daily life tasks. Since Instagram features are frequently used for daily communication needs in peer-to-peer chat or in a small group, it is worth exploring the use of these features within a task-based language teaching (TBLT) framework. Integrating CMC with TBLT promotes language learning through authentic, unrehearsed, meaningful real-life tasks (Ellis et al., 2019). However, utilizing digital technology in language instruction without research-informed learning models might reduce the possibility for English language learners to reach optimum learning outcomes (González-Lloret & Ortega, 2014). Thus, this study married TBLT with CMC as an online learning instruction model. This present study sought to explore students' experiences using multimodal CMC tasks on Instagram and determine how this online learning experience could foster their English language communication skills in an Indonesian context. The research questions (RQs) were:

- What are students' perceptions of completing English language learning tasks through CMC?
- What are the affordances of the text, audio, and video features of Instagram chat as the CMC platform for task completion?
- What are students' perceptions of the use of nonverbal cues (e.g., emojis, GIFs, images, gestures) for fostering English communication?
- What are examples of nonverbal cues used in the text and video chat for fostering English communication?

2. Literature review

2.1. The role of CMC in language learning

Studies suggest that CMC holds great potential for additional language (henceforth LX, see Dewaele, 2017) use and learning (Castrillo, 2013). For example, discussions with classmates conducted through Instagram comments about the content posted have been shown to stimulate students to produce more language output, thus helping them enhance their communication skills (Aloraini, 2018). In addition, it also influences students' affective states, such as motivation when interacting with an authentic audience while doing real-life language learning tasks (Chen & Brown, 2012). It might also reduce students' anxiety due to CMC being a digital space in which direct face-to-face interaction is unnecessary (Côté & Gaffney, 2018).

With the synchronous function, SCMC enables students to connect with others in real-time virtually and supports spontaneous conversation, which may improve students' communication skills (Huang, 2018). This is because SCMC peer-to-peer interactions effectively enable students to interact and practice target language outside the face-to-face classroom (Lin, 2015). Additionally, the multimodality available in SCMC, such as live text chat, audio, and video call, offers optimal alternatives for students to strategically select their preferred channels following their digital learning styles and conversation needs (Guichon & Cohen, 2016). As a result, SCMC provides students with an authentic way of language learning, with varieties of channel options that are directly connected to real-life communication needs.

2.2. The synergy between TBLT and CMC

Task-based instruction is an approach that requires the implementation of communicative tasks to develop learners' communicative competence, problem-solving, and collaboration skills. The emphasis is on learning by doing, with a focus on the use of authentic language. According to Ellis (2018), four key principles undergird the nature of a task: (1) the primary focus during communication must be meaning, (2) there must be some kind of knowledge gap, (3) there must be a communicative outcome, and (4) learners should be encouraged to tap into both linguistic and non-linguistic resources to complete the task. In other words, learners can draw upon their existing LX linguistic repertoires, such as grammar and vocabulary, as well as L1 knowledge, and non-linguistic resources, such as gestures and facial expressions, to facilitate task-based interaction. Thus, the task can stimulate interactions in which learners may experience the dynamic relationship between language knowledge and language use in a particular social context (Samuda & Bygate, 2008). Based on the kind of gaps in the task (Principle 2), a task can be classified into three types (Ellis, 2018; Prabhu, 1987): *information gap* (students have to provide the missing information by interacting with each other), *reasoning gap* (students draw new information through a process of inference, deduction, or practical reasoning), and *opinion gap* (students identify and articulate their personal preferences, feelings, or attitude to complete the task).

For decades, researchers have examined the interrelationships between CMC and task-based instruction as a framework for language learning (Ziegler, 2016), whether as informal activities for sustaining interaction and learning in an outside classroom context, or as a pedagogical framework for designing effective instructional activities in online language courses. González-Lloret and Ortega (2014) assert that CMC offers a more optimal environment for students to learn an LX and improve their digital literacy and skills than a traditional classroom. Levak and Son (2017), for example, noted that the use of Second Life and Skype developed students' listening comprehension skills and boosted their motivation. Students were also reported to have a better grasp of intercultural schema after interacting with their international counterparts. In addition, Castañeda (2019) highlighted the benefits of VoiceThread for improving students' productive and receptive communication skills.

Furthermore, some studies have reported on the effect of CMC modality in conjunction with specific task types on learners' communication skills. For example, the implementation of jigsaw and decision-making tasks in text-based interaction through internet relay chat (IRC) was found to stimulate more negotiation of meaning and awareness of strategy use on self-repair, asking for clarification, and codeswitching, during a communication breakdown (Kost, 2008). Audio-based interaction in a 3D virtual world of Second Life with jigsaw and information gap tasks was reported to promote more negotiation of meaning and communication strategies than did decision-making and opinion exchange (Chen, 2018). The integration of video-based interaction using Skype and narrative tasks was proven to facilitate students' language development as effectively as face-to-face interaction conducted in a physical classroom (Rassaei, 2017).

2.3. Research on multimodality and CMC tasks in online language learning

Multimodality in the context of language learning often refers to the coordination of multiple different modes of communication, both verbal and nonverbal, to deliver a united meaning (Dressman, 2020). Rather than relying on one single mode of communication, people also tap into multiple semiotic resources, such as speech, text, images, and gestures, to construct meaning (Jewitt, 2014). Generally, while people may assume that verbal utterances are the most informative content available to send a message when communicating via the internet, nonverbal elements (e.g., gestures, eye gaze, or images) also contribute significantly to constructing meaning in communication (Jewitt, 2014; Norris, 2004). Hence, verbal and nonverbal communications are integral parts of human interaction and equally add to the intended meaning.

Due to its communicative goal, task-based instruction respects the spontaneous use of both nonverbal and verbal cues for completing CMC tasks. Nonverbal cues serve to convey a meaningful interaction in a language learning context (Dressman, 2020). These features shape the way we communicate and produce multimodal language outputs that are playful, creative, and spontaneous. Prior studies also examined the influence of multimodality on LX communication, with special attention paid to the role of different forms of nonverbal cues to shed light on how students utilize these in online communication. The use of gestures in a videoconferencing task, for example, may extend the negotiation of meaning discourse by constructing vocabulary meaning with the aid of gestural moves (Lee et al., 2019). Another nonverbal resource, gaze, was also reported to facilitate the establishment of social presence during online open-ended task communication (Satar, 2013).

Previous CALL research has asserted that nonverbal cues have a distinctive function related to a particular CMC channel. For example, text-based CMC emojis serve as a replacement for the absence of nonverbal cues used in face-to-face communication and can function pragmatically to signal emotions, backchannel devices, humor, or irony (Li & Yang, 2018). Emojis may also operate as communication strategies to resolve the problem of showing a dramatic expression in written messages, such as an emoji face screaming in fear 😱 for extreme feelings of dread (Hung & Higgins, 2015). Moreover, video-based CMC embodies both verbal and visual modes of communication; the latter includes gestures that reinforce the negotiation of lexical meaning for interlocutors to convey and better comprehend the intended meaning (Lee et al., 2019). Facial expression and gestures also support students' socio-affective communication needs and improve fluency (Satar, 2016). Although the reviewed literature above has investigated the role of nonverbal cues in CMC tasks, further research comparing students' experiences while using nonverbal cues to undertake communicative tasks across communication channels (e.g., text-chat, audio-chat, and video-chat) is relatively underexplored.

3. Methodology

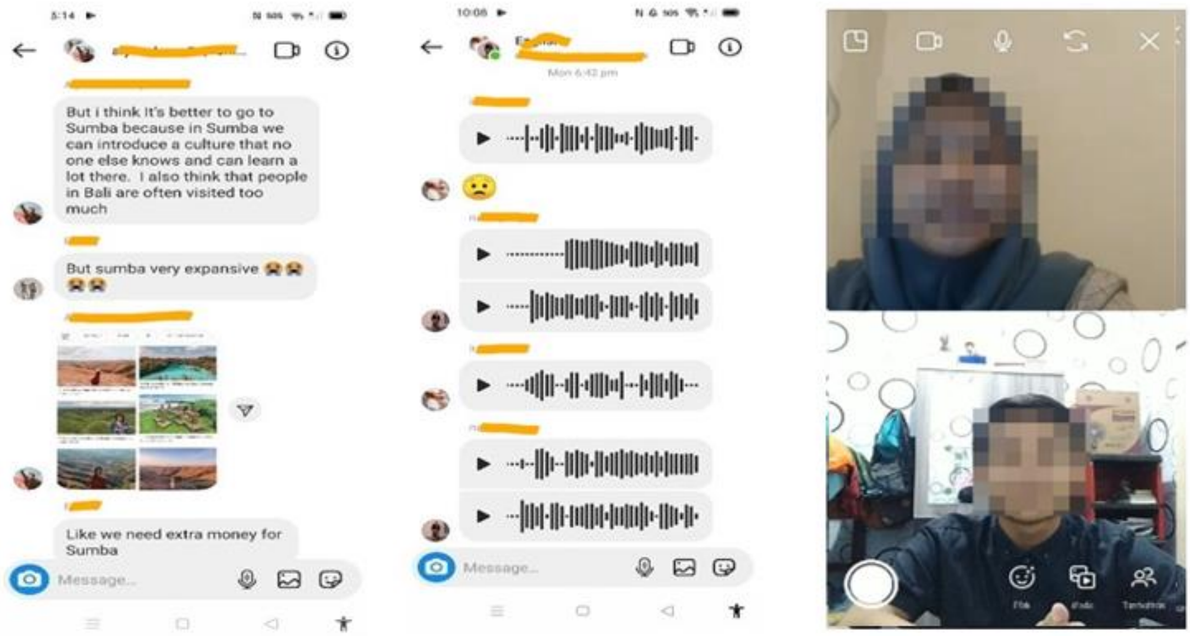
3.1. Participants and the context

The pilot study described here involved 30 first-year business school students (F = 22, M = 8) enrolled in the General English unit at a university in Indonesia. The unit was run online through a massive open online course (MOOC) “OpenLearning platform” due to the COVID-19 pandemic. It was taught based on the module mandated by the university. Since Indonesian students must pass the university entrance examination before being enrolled in a university, the students’ English language proficiency was assessed as ranging between elementary and lower intermediate level based on the Common European Framework of Reference (CEFR). All students were aged between 18 and 21 years (M = 18.46, SD = 0.86). Participants were recruited from three different formal online classes. They had not met in person before. They were invited to participate in the study and voluntarily signed up. The recruitment principle ensured that all participants met particular criteria, such as being familiar with the Instagram application and having adequate smartphone hardware to access all three channels, text-chat, audio-chat, and video-chat. Human ethics approval was granted for the study, and students further signed a consent form stating that they agreed to have their photographs made public and published. Before conducting one-on-one semi-structured interviews, the participants’ communication performance was measured and classified into high, medium, and low achievement, based on their language production and motivation during task completion. It was done by counting the number of turns they took. Further, three participants from each group were randomly selected for interviews using a voice call through WhatsApp as the voice call feature was not available on Instagram at that moment. The interviews were digitally recorded for data analysis.

3.2. Instagram as the CMC platform

This study implemented Instagram as a CMC platform based on the following rationale. First, Instagram allows users to freely utilize multimodal features for communication via either their mobile phone or a web browser. In Indonesia, Instagram was ranked as one of the top three SNSs, followed by YouTube and WhatsApp, in 2020 (Nurhayati-Wolff, 2021a). Around 70% of Instagram users were aged between 18 and 34 years. According to the needs analysis done in the study, Instagram was participants’ preferred mobile-based CMC for language learning. One of its salient affordances enables users to select various types of communication channels (e.g., text-chat, voice-chat, and video-chat, as seen in Figure 1) or flexibly combine them with other dynamic features, such as emojis, GIFs, images, and filters.

Figure 1. Screenshots of Instagram chat interface (left to right: text-chat, audio-chat, and video-chat)



3.3. Procedures

Since the students were expected to be involved in authentic, contextual, and functional use of language through CMC tasks, three types of communicative task (e.g., information gap, reasoning gap, and opinion gap) were designed and implemented in this study, following the task typology suggested by Prabhu (1987) and Ellis (2018). The task contents were developed in line with the task principles of focusing on meaning, having some kind of knowledge gap, producing communicative outcomes, promoting both linguistic and non-linguistic resources to complete the task, and real-world task resemblance (Long, 2015). Before CMC task implementation, a needs analysis was conducted to gather information on the students' communication needs and preferred task topics. Each task was performed twice (except the opinion gap task that ran three times); the first time, students were in a dyad, and the second time, they were in a group of three. The participants were instructed to use a specific channel in the first iteration (e.g., text, audio, or video chat). In contrast, they could select their preferred channel to communicate (free channel) in the second iteration. We considered the nature of each task concerning the functionality of a communication channel before assigning the tasks. For example, students were requested to use video calls in Tasks 1 and 7 that required visual aids (video and pictures). Text chat was used in Task 3, which encouraged students to create a shorter and more efficient exchange of information. Audio chat was selected for Task 5, supporting the participants to express a complex opinion. Furthermore, we allowed free channels (Task 2, 4, and 6) for group activities to explore the possibility of using mixed Instagram channels to complete the tasks (see Table 1 for the summary).

Table 1. Summary of the task types and instruction

	Task types and instruction
Week 1	Information gap; video chat; in a dyad
	Rearranging random short videos become a full story (20-30 minutes)
Week 2	Information gap; free channel; in group
	Rearranging jumbled pictures become a complete story (20-30 minutes)
Week 3	Reasoning gap; text chat; in a dyad
	Discussing to select tourism destination sites for four days' holiday (20-30 minutes)
Week 4	Reasoning gap; free channel; in group
	Selecting only 12 kg survival kits from the provided list to carry on during the journey (20-30 minutes)
Week 5	Opinion gap; audio chat; in a dyad
	Sharing and discussing opinions about 'how to build a strong friendship?' (20-30 minutes)
Week 6	Opinion gap; free channel; in group
	commenting, sharing, and discussing two pictures showing contrast life phenomena (happy and sad family pictures) (20-30 minutes)
Week 7	Information gap; video chat; in a dyad
	Guessing six different characters taken from famous novels and movies (20-30 minutes)

3.4. Data collection methods

We employed a mixed methods research design to obtain in-depth information from the collected dataset and allow for data triangulation to ensure the validity and reliability of the quantitative findings and the trustworthiness of the qualitative findings (Creswell, 2012). The quantitative and qualitative data were gathered from various sources, including pre-and post-study surveys, students' reflection journals, and semi-structured interviews. The surveys were created in Qualtrics and consisted of closed-ended items scored on a five-point Likert scale ranging from *strongly disagree* to *strongly agree* (Link of pre-study survey https://curtin.au1.qualtrics.com/jfe/form/SV_2sfKSwSzNqUvkUd and post-study survey https://curtin.au1.qualtrics.com/jfe/form/SV_3qRuwCMlnfTSDWJ) The surveys were used to acquire information about students' perceptions regarding tasks, Instagram features, and the perceived benefits on their English language proficiency of using multimodal CMC tasks. The items on pre-and post-study surveys were identical as we intended to compare the experience before and after the study. The survey was developed by adapting the items from prior CMC and SNSs research (e.g., Aloraini & Cardoso, 2020; Erarslan, 2019; Lee & Markey, 2014), targeting key constructs, such as affective aspects and perceived benefits related to the use of CMC for language learning. Further, the items were reviewed by two TESOL/Applied Linguistics specialists, and member checked with the teachers to clarify ambiguous items (Dörnyei & Csizér, 2012). We also utilized guiding prompts to elicit information for students' reflection journals. Three high, medium and low achievers (N = 9) were interviewed to gather more information about factors affecting their task completion, the challenges in performing tasks, preferences for task types, and experiences with using multimodal features to complete the tasks. The survey items, journal prompts, and interview questions were all in Bahasa Indonesia to reduce any misinterpretations by participants due to their English language proficiency (Supplementary materials. To view survey items, journal prompts, and interview questions referred to in this article, please visit https://bit.ly/3GBXFna).

This study was conducted over nine weeks. Participants were asked to take the pre-study online survey using a shared link in the first week. Then, the participants were required to select the preferred time to take part in task implementation, as they studied online from home and lived in different geographical regions around Indonesia. In the following seven weeks, participants performed the task-based instruction using their smartphone and Instagram account with guidance from one of the researchers. The researcher's role was participant-observer and involved organizing the online meetings, delivering the task materials, assigning the dyads or groups, and observing, monitoring, and recording each ongoing task session using a smartphone. First, an Instagram whole group chat was created as the main communication channel between the researcher and participants. The researcher posted task instructions and a list of dyads or groups of participants. To form dyads or groups in each task, the participants were randomly selected. In performing the tasks, the participants created a small group chat consisting of the designated students and the researcher. Hence, the researcher was able to monitor students' task interactions and document the process. After finishing each task, participants also kept a journal reflecting on their experiences with Instagram CMC tasks. They considered their experiences of, and attitudes towards, doing the CMC tasks, guided by the reflection prompts. At the end of the study, participants were invited to take the post-study survey, and nine students were interviewed.

3.5. Data analysis methods

Data were analyzed using a convergent mixed method design in which quantitative and qualitative data were analyzed separately, and then the results were compared to see whether they ratified each other (Creswell & Creswell, 2018). First, the qualitative data gathered from the reflection journal and interviews were translated into English and analyzed thematically (Miles et al., 2014). For the thematic analysis, each researcher read students' reflections and interview responses repeatedly and coded the units derived inductively across the qualitative dataset. We then discussed the coding discrepancies and reached a consensus in resolving them. These units were compiled into groups of specific patterns, and then the patterns were compared and classified into several categories. All similar categories were then subsumed into higher-level themes to capture the essence of the participants' perceptions. Meanwhile, the Likert scale survey responses were quantitatively measured to identify trends in students' perceptions before and after the implementation of CMC tasks, the affordances of Instagram features, and students' use of nonverbal cues. Further, we compared the mean scores of survey results using a paired *t*-test via SPSS 22.0. Before running the paired *t*-test, the data had been measured as normally distributed.

4. Findings

4.1. Quantitative results

The quantitative results from the pre-and post-study surveys addressed all RQs in three parts: (1) the use of CMC tasks, (2) the use of multimodal features of Instagram, and (3) the use of nonverbal cues for online communication (see Tables 2, 4 and 5).

4.1.1. The use of CMC tasks

Table 2 presents the results obtained from the pre-and post-study surveys related to students' perceptions of using tasks to foster English communication. It can be observed that the students positively perceived the tasks to facilitate English use, meaningful and prolonged communication (96%; Q3), and improved collaboration with peers (from 44% in the pre-study to 96% in the post-study) (Q2). Their motivation for communication in English (Q6) also increased from 79% in the pre-survey to 90% in the post-survey, and more than 96% confirmed that they were actively engaged in tasks (Q1).

Overall, students' perceptions regarding the use of tasks positively changed. The results for the post-study survey (M = 4.41, SD = 0.33) were significantly higher than the pre-study survey (M = 3.77, SD = 0.33). The paired *t*-test found a statistically significant difference in student perceptions of using tasks to foster English communication before and after completing the tasks ($t(29) = -6.38$, $p < .001$, $d = 1.2$) (see Table 3).

Table 2. Students' perception of using CMC tasks

Questions of the survey		*SD*	D	N	A	SA
1. I participated actively in English learning tasks (e.g., discussion for deciding holidays destination with partner) through Instagram.	Pre Post	0.0% 3.3%	3.4% 0.0%	17.3% 0.0%	65.5% 30.0%	13.8% 66.7%
2. I worked cooperatively with others both in pairs and in groups during English learning tasks through Instagram.	Pre Post	6.9% 3.3%	10.4% 0.0%	37.9% 0.0%	41.4% 36.7%	3.4% 60.0%
3. I felt that working with peers when doing tasks through Instagram helped me better communicate in English.	Pre Post	0.0% 3.3%	0.0% 3.3%	13.8% 0.0%	75.8% 30.0%	10.4% 63.3%
4. If I encountered difficulties, I still tried my best to complete those tasks.	Pre Post	0.0% 3.3%	0.0% 0.0%	6.9% 0.0%	62.1% 30.0%	31.0% 66.7%
5. Doing tasks through Instagram motivated me more to communicate in English.	Pre Post	0.0% 3.3%	0.0% 0.0%	20.7% 6.7%	58.6% 53.3%	20.7% 36.7%
6. I enjoyed working with peers when completing tasks through Instagram.	Pre Post	0.0% 3.3%	0.0% 0.0%	17.3% 0.0%	69.0% 53.3%	13.8% 43.4%

Note. SD: Strongly disagree, D: Disagree, N: Neutral, A: Agree, SA: Strongly agree.

Table 3. Students' perception of using CMC tasks

The use of CMC tasks	*n*	Mean	*SD*	*t*	*p*
Pre-survey	30	3.77	0.37	-6.38	.001*
Post-survey	30	4.41	0.33		

Note. $^{*}p < .05$.

4.1.2. The use of multimodal features of Instagram

Table 4 shows the students' responses to using the multimodal Instagram features to perform CMC tasks. Overall, students agreed that Instagram multimodal features rendered enjoyment (93%; Q7), eased anxiety in English communication (76%; Q8), and provided flexibility in communication modes (93%; Q9).

Table 4. Students' perception of using multimodal Instagram features

Questions of the survey		*SD*	D	N	A	SA
7. I enjoyed using Instagram features (e.g., text, voice, video, images) to communicate in English.	Pre Post	0.0% 3.3%	0.0% 0.0%	27.6% 3.3%	48.3% 40.0%	24.1% 53.4%
8. I felt more comfortable communicating in English using different Instagram features than only using one mode.	Pre Post	0.0% 3.3%	3.4% 3.3%	31.1% 16.7%	41.4% 56.7%	24.1% 20.0%
9. Instagram features facilitated me with more options to communicate in English.	Pre Post	0.0% 3.3%	0.0% 0.0%	24.1% 3.3%	51.8% 56.7%	24.1% 36.7%

Note. SD: Strongly disagree, D: Disagree, N: Neutral, A: Agree, SA: Strongly agree.

Students were generally satisfied with the additional CMC options that the multimodality of Instagram offered to further support their English communication. Table 5 demonstrates a significant difference in the students' perceptions of using the multimodal features between the pre-study survey (M = 3.94, SD = 0.66) and post-study survey (M = 4.27, SD = 0.46). The paired t-test results showed ($t(29) = -2.20$, p = .036, d = 0.3); this meant that students acknowledged the positive effect of multimodal Instagram features on enhancing English communication.

Table 5. Students' perception of using multimodal Instagram features

The use of Instagram features	*n*	Mean	*SD*	*t*	*p*
Pre-survey	30	3.94	0.66	-2.20	0.36
Post-survey	30	4.27	0.46		

4.1.3. The use of nonverbal cues

Table 6 shows the students' perceptions of using nonverbal cues for online communication. Multimodality was recognized to be important for English communication (96%; Q13). Students also highly valued nonverbal features (e.g., emoji, GIFs; 93%; Q14) and the use of gestures in video chat for English communication (93%; Q15).

Table 6. Students' perception of using nonverbal cues

Questions of the survey		*SD*	D	N	A	SA
10. Using different modes (e.g., text, voice, video, body language, visual, images) was important for English communication.	Pre	0.0%	0.0%	13.8%	55.2%	31.0%
	Post	3.3%	0.0%	0.0%	53.4%	43.3%
11. Nonverbals features (e.g., emojis, GIFs) of Instagram were valuable for English communication.	Pre	0.0%	3.4%	27.6%	48.3%	20.7%
	Post	3.3%	3.3%	0.0%	46.7%	46.7%
12. Gestures through Instagram video chat were valuable for English communication.	Pre	0.0%	0.0%	37.9%	48.3%	13.8%
	Post	3.3%	0.0%	3.3%	60.0%	33.4%

Note. SD: Strongly disagree, D: Disagree, N: Neutral, A: Agree, SA: Strongly agree.

Overall, as illustrated in Table 7, using nonverbal cues for English communication was well received by the students. The results of the post-study survey (M = 4.39, SD = 0.45) were higher than the pre-study survey (M = 3.93, SD = 0.62), and the paired t-test results showed a statistically significant difference in the students' perceptions of using the nonverbal cues, indicating that students acknowledged the positive role of multimodal features in fostering English communication ($t(29) = -2.90$, $p = .007$, $d = 0.5$).

Table 7. Students' perception of using nonverbal cues

The use of nonverbal cues	*n*	Mean	*SD*	*t*	*p*
Pre-survey	30	3.93	0.62	-2.90	$.007^{*}$
Post-survey	30	4.39	0.45		

Note. $^{*}p < .05$.

4.2. Qualitative findings

4.2.1. Students' perceptions of using CMC tasks

To answer RQs 1 and 2, the emerging themes from the students' journal reflections and interviews regarding their perception of using CMC tasks for fostering English communication were analyzed. Finally, the challenges they identified in performing tasks were included at the end. All names displayed are pseudonymous.

Supporting LX production. Many students perceived participating in CMC tasks as facilitating their target language production. The CMC tasks themselves seemed to have enabled them to use more LX in their daily communication outside the classroom since they had limited opportunities to practice during the formal lessons (Vandergriff, 2016). Having tasks through CMC created an optimal virtual space and stimulated students to use the target language in an authentic situation since the tasks were designed to resemble real-life communication (Aloraini, 2018). Putri identified her experiences as follows:

In Instagram, I straightaway practiced, . . . In the classroom, we were usually given (teaching) material first. It means we listened to the lecturers talking and we didn't always practice. (Interview/Putri)

In addition, Instagram features were reported to create more flexible communication. Students appreciated the interchangeability between text-chat and voice note features. This could be seen from one student's testimony:

There were times when I did not know the pronunciation of a word in English, so I chose to type it. On the other hand, when I had to explain something long, I preferred to use a voice note instead of having to type it. (Journal/Mukti)

With the user-friendly interface, students just needed to touch the record button to operate voice notes, speak as needed, and then send the recording to their partners. Interestingly, although many students reported practicing CMC tasks with partners supporting LX production, some students complained that not all partners could work with them to complete the task. One student (Adi) observed that sometimes there was a group member who tended to be silent or unresponsive in the group talk. As he stated, "Thank God, I got a partner who was equally active too, we were also good, we were not nervous."

Encouraging self-confidence and motivation to learn LX. Another positive finding was that multimodal CMC tasks boosted students' agency and motivation to use the LX. As the study progressed, some students gradually developed a sense of self-confidence for communication in English. The exposure they got from exchanging information during CMC tasks was considered to aid in language intake that could serve as a confidence booster. CMC tasks reinforced students' confidence to communicate in English outside of the classroom (González-Lloret & Ortega, 2014). Zahra voiced:

When I did that at first, I did not feel confident because I was still confused with English . . . but more and more I got used to (with the tasks by) finding vocabulary, the right sentences, and how to apply it correctly. (Interview/Zahra)

Rudi mentioned that the use of Instagram as a platform enhanced his self-confidence. Its facility in delivering online communication without physical meetings reduced anxiety:

With the help of social media like Instagram, it could help me increase my confidence because we did not face (to-face) directly with others. (Interview/Rudi)

In terms of learner motivation, many students appreciated the CMC tasks that helped them develop a sense of learning autonomy. Due to their involvement in completing the tasks, students were motivated to access online materials from various sources, such as streaming videos, to acquire more vocabulary. For example, Alevi confessed, "I realize that sometimes I watch something (videos) in English to improve my English vocabulary." They also reviewed the recorded video of their interactions with partners to further scaffold their speaking skills and pronunciation.

Providing enjoyment and fun. Many students perceived that the tasks were linked to their communication needs in daily life, such as deciding where to go, sharing opinions related to a specific phenomenon, or exchanging information about interests. The CMC tasks were considered to be dynamic learning activities as vouched for by Adi: "because there was interaction, there was discussion, and there was debate too." The use of Instagram as a medium was considered to provide enjoyment in two aspects. First, it served as a vibrant communication tool for language learning, as the students felt that visual features such as emojis and GIFs could mediate discourse and add another layer of expression to make their communication more effective and enjoyable. As Irina expressed, "I enjoyed it . . . I could also post gifs, uhm . . . pictures that moved as if I felt what was represented in the emoji on Instagram." Second, with Instagram's basic function for establishing social networking, students were also delighted to gain many new friends on Instagram. Being able to share an opinion and work together with new friends while completing the tasks was fun for them since they had limited access to socializing with their colleagues due to social distancing during the COVID-19 pandemic. Through their involvement in the CMC tasks, they were able to mingle with their peers from other classes and increase the numbers of their Instagram followers as well.

Improving digital literacy skills. Some students felt that doing tasks through Instagram helped them acquire the digital skills they needed. As technology has developed so quickly, they commented on not being able to keep up. They lacked knowledge about the current features and facilities offered by the internet (e.g., cross-communication channels, the plethora of language aid applications, and visual features). By practicing English using Instagram tasks, students had to explore non-linguistic resources, including the availability of multimodal channels and other online tools as useful alternatives to help them complete the tasks (e.g., video calls, internet search engines, and online dictionaries). As indicated by Zahra, "I just found out that there was a video call on Instagram" (journal), and Irina also sought additional information from the internet: "When I didn't know how to express it in English, I would immediately search (meaning in internet browser) for it" (interview). Students also became aware of the affordances and constraints of each Instagram channel for communication, as suggested by Haque: "Text chat is a simple but very important feature; in this case, voice chat is to clarify the pronunciation, and video calls are more effective in discussions" (journal).

Challenges in implementing the multimodal Instagram tasks. Some challenges were also reported during student performance of the CMC tasks. First, the internet connections around their locality were sometimes not stable, thus obstructing the completion of the task process:

Sometimes the signal went out, as during a video call, the screen was blurry, the picture was paused, the sound was sometimes unclear, so it was a bit annoying. (Interview/Putri)

As Indonesia is an archipelago country, internet coverage is not equally distributed, especially in rural areas. The signal quality was poor and insufficient to transfer both audio and visual data requiring higher bandwidth. Another student, Feby, also commented that, “If the signal was not good, we could not do the task well.” This finding indicates the importance of internet stability, which was integral to any type of CMC task and its completion.

Second, most students felt very anxious when doing CMC tasks, given the fact that their English class rarely used CMC tasks as a learning platform. As Feby stated:

In the first task, I was inexperienced . . . so, I felt a little scared, confused, afraid of being wrong, and worried whether the pronunciation was correct or not. (Interview/Feby)

These initial qualms and anxieties about performing CMC tasks arose because the students had little experience of performing tasks through CMC. These tasks required them to actively produce language output and hone their communicative skills through speaking and writing spontaneously with their classmates. Students were sometimes not ready for the shift from receptive to productive-based tasks, especially through SCMC. This seemed to be compounded by their English language proficiency level. Satar and Ozdener (2008) suggested that text chat was appropriate for less proficient or more anxious students, while video chat was suitable for more proficient and less anxious students. However, students’ anxiety in our study was reduced as they performed the tasks weekly, supporting the claim that students’ anxiety will gradually decline once they have achieved familiarity with tasks settings (Gurzynski-Weiss & Baralt, 2013).

4.2.2. Students’ perceptions toward the use of Instagram features and nonverbal cues

We also compared the students’ perceptions of multimodal communication enacted in the three main types of Instagram channels: text-chat, video-chat, and audio-chat, and provided the example of nonverbal cues used in text and video chat for addressing RQs 3 and 4.

Text-chat feature. Most students felt that multimodal features of text-chat, such as images, emojis, and GIFs, helped them express and convey meaning in various ways:

When I didn’t understand a word (related to a feeling) in English, I could express it through emojis or GIFs. (Journal/Andria)

Images could provide a visualization when I imagined a place that I had not visited yet. (Journal/Ferry)

Figure 2. The use of emojis, GIFs, and images in a text-chat interaction

The excerpts above show the value of emojis, GIFs, and images replacing or strengthening verbal messages in online interaction. It can be seen from the conversation in Figure 2 how students employed these visual features during Tasks 3 and 6. For example, when students had to give their opinions about two contrasting pictures, one which illustrated harmonious families and one which showed discordant families in Task 6, they used emojis (A) to express sadness and empathy by adding "a worried face emoji." A GIF (B) was also added to convey a more vivid feeling of sadness through an animation representing a crying girl. They also used the actual image of the object "a water palace" (C) to back up the verbal message sent earlier when discussing tourist destinations in Task 3. Thus, the combination of verbal and nonverbal channels livened up students' use of the English language.

Audio-chat feature. Many students found using voice-chat beneficial for carrying out the tasks. They positively perceived the functionality of voice-chat in bringing suprasegmental elements (e.g., intonation and stress) into the message, thus helping the interlocutors to interpret their words correctly. As Fiona stated, "Using the voice chat feature, I could also bring feelings directly into the sentence, which was interpreted in the form of intonation." In addition, the voice note function in Instagram was user-friendly, allowing students to prepare what they would like to say and replay by rechecking the recording for pronunciation and grammar. When students made a mistake, they could erase it and repeat the recording before it was sent. As noted by Irina:

> For voice notes, it was easy because, for example, when I felt wrong in recording a voice note, I just slid it and deleted it and repeated it. (Interview/Irina)

Video-chat feature. Most students also acknowledged that non-verbal cues, such as gestures and facial expressions, were conducive to online video conversation. They reported that conveying a message with gestures could help the interlocutor to better understand since gestures could carry referential (semantic) meaning:

> Gestures are very helpful, especially when speaking in English; it's like when someone didn't know what I meant, I could use gestures. (Interview/Rani)

Students used gestures, gaze, and facial expressions for online interaction during Task 7 (information gap). Examples of gestures used in online interactions through video calls on Instagram are shown in images in Table 8 below. Here, Irina (IR) described one of the pictures using verbal and nonverbal cues. She used a referential gesture (a gesture that resembles aspects of the semantics and lexical content) (Kendon, 2004) by moving her hand to the back of her shoulder and pointing to the position of the wing (A) while saying, "the character has a . . . black cloth and *sayap* (wing)." It appeared that the gestures emerged when she attempted to find an appropriate word for saying "wing" but failed to retrieve the word, whereupon she used a translanguaging strategy where she called upon the word *sayap* in Indonesian. In this excerpt, referential gestures were used to direct attention to the specific objects that served as distinctive characteristics of the superhero described. They cognitively supported the lexical search process. For the second gesture, Irina directed her index finger to the space in front of her face while saying "black mask" (B). This gesture acted as a co-expressive function since it seemed to be formed at the same time as uttering the phrase "black mask" and indicated the real object worn on the face (Negueruela & Lantolf, 2008). Therefore, it seemed to provide additional information about the type and position of the mask. For the gaze and facial expression, Irina looked at the screen while describing the character with a friendly and relaxed expression, while Askia (AS) paid serious attention by listening and watching Irina's explanation. Askia changed her facial expression when trying to guess the character's name. Hence, the gaze and facial expression built social presence and emotional connectivity between the interlocutors (Satar, 2013).

Table 8. The use of gestures in video call interaction

Speech	Scenes	(I) Gaze and facial expression	(II) Gestures
IR: Ok . . uhm . . next . . uhm, the black mask . . character has a . . . black cloth or *sayap eh apa* (wing eh what) in on the back . . . black mask AS: Oh batman, I know this. ((laugh)) IR: Uhm, yeah, good answer	(IR vs. As)   A B	IR: Looks at the screen while describing the character AS: Seriously pays attention while watching IR's explanation, and she changes her facial expression when she realized the character's name	IR: She moves her hand to the back of her shoulder to point to the position of the character's wing (A) and she also tries to direct her index finger in front of her face to indicate "mask" (B)

5. Discussion

The first RQ explored students' perceptions toward completing English language learning tasks through CMC. Quantitatively, students reported positive experiences with the implementation of multimodal CMC tasks, as shown by the statistically significant results between the pre-and post-surveys, indicating that the multimodal CMC tasks were beneficial for fostering English communication (Table 3). Based on the qualitative findings, students characterized multimodal CMC tasks as promising mediums for practicing all language skills in an authentic setting. In line with the previous research findings (Aloraini, 2018; Vandergriff, 2016), CMC tasks in this study encouraged students to immerse themselves in the LX environment more deeply than in the classroom. In the CMC environment, they could spontaneously use more target language in an authentic setting. Students recounted that their involvement in CMC also created a space for them to use authentic target language for their real-life communication needs by conversing and interacting with their friends from different language backgrounds. This finding resonated with Erarslan's (2019) report that Instagram enhanced students' use of the LX as well as focusing on collaboration, cooperation, and sharing knowledge among the users.

Regarding students' affective feelings, the findings showed that the CMC tasks were pleasant and motivating activities. Students respected the familiarity of topics discussed in the tasks, which allowed them to connect the contents to their personal lives. This finding supports Chen and Brown's (2012) claim that performing life-like tasks can stimulate students to produce more meaningful and accurate LX. Besides this, the informality of interaction provided by CMC tasks can bring about freedom for students to express their ideas in the target language without being limited to the scripted dialogue often given in Indonesian classrooms. It echoes Lee's (2016) conclusion that open-ended CMC tasks provide freedom for discussing and understanding the topic. Some positive points related to CMC use were also noted by the students, such as reinforcing their confidence, reducing their anxiety, and offering enjoyment. These positive affective experiences are key to maximizing the potential for language learning by creating a positive emotional bond to learning experiences (Dewaele & MacIntyre, 2014) and intensifying students' participation and engagement in LX learning (González-Lloret & Ortega, 2014).

The second RQ sought to identify the affordances of the text, audio, and video features of Instagram as the CMC platform for task completion. A statistical majority of students found that multimodal Instagram features afforded LX communication and made it enjoyable, comfortable, and flexible, which helps LX production (Table 5). The students shifted between multiple modes, increasing opportunities for interaction and complementing each other's modes of communication. The findings evidenced Hampel and Stickler's (2012) statement that CMC multimodal features can be used to enhance language teaching and learning. However, some drawbacks were also documented. The first was internet infrastructure. Many students involved in this study lived in rural areas and often experienced poor internet connections. Thus, in implementing CMC tasks, particularly via video call, language instructors should consider the stability of the internet networks in their locality to ensure that all the channels will operate well. Second, although students had been familiar with Instagram, they still felt very anxious at the beginning of task implementation, especially in completing tasks through voice chat and video chat since they had rarely practiced English speaking through these channels. Our study also confirmed Satar and Ozdener's (2008) claim that less proficient LX users feel more anxious using voice and video chat compared to text chat. Any use of CMC as a language tool would need to consider this and be sure to familiarize students with the tasks by grading the level of task complexity and staging the channels from text chat to voice chat and then, lastly, video chat.

RQs three and four concerned the students' experiences and perceptions of using nonverbal cues (e.g., emojis, Gifs, images, intonation, gestures) in three different Instagram communication channels (text, audio, and video chats) for their English communication and examples of the nonverbal cues used. We presented nonverbal use focusing only on text and video chat due to the space limitations in this article. The affordance of CMC to offer nonverbal features in online communication provided a vibrant atmosphere to the interaction. Students acknowledged nonverbal cues as an essential tool for interpreting meaning in LX communication since the communication was naturally multimodal. They underscored the idea that semiotic resources (co)constructed meaning and contributed significantly to the determination of meaning in communication (Jewitt, 2014). Emojis and other visual cues encouraged colorful and socially bound interaction. Similar to Chang's (2016) finding that emojis possibly made the interaction more socio-emotional and less face-threatening, students also found that the use of emojis, GIFs, and images facilitated a communication strategy in the absence of visual cues in written based communication, as noted in Hung and Higgins's (2015) study.

Referential gestures were often used by students with inadequate LX vocabulary resources for lexical search functions (Table 8, Image A). Gestures were formed before the students had found the appropriate words to convey the meaning. During this process, students commonly continued to speak with a filler such as "uhm" (Lee et al., 2019; Negueruela & Lantolf, 2008). Additionally, the referential gesture served as a spontaneous co-expressive function to accompany speech (Table 8, Image B). It seemed to contribute additional information about the real object described. These gestures also acted as a strategy to elucidate meaning for the interlocutor by aligning with the language used to create a coherent semantic message and enabling dual-channel input. This was important for language learning. However, the use of nonverbal cues did not always result in successful communication due to different interpretations from students with different cultural backgrounds. This had the potential to lead to misunderstandings.

6. Conclusion

The findings of this study suggest that Instagram as a means of CMC could be a conducive medium for Indonesian students to practice their English language authentically. It could also be a vehicle for socializing and connecting people to the global community. Despite the constraints found in the study, Instagram still proved to be a valuable tool for fostering language learners' communicative skills beyond the classroom walls. Further, the analysis of student use of multimodality features added to our understanding of how learners engage with verbal and nonverbal semiotic resources within various communication channels and their interrelationship with task-based instruction for enhancing English communication. Thus, it is essential to raise awareness of teachers and students to this and explicitly promote the value of nonverbal cues as well as verbal cues for LX communication. It has to be remembered, however, that this study was limited to one group of students in an Indonesian context. Nonetheless, given the growth of multimodal language learning materials and teaching resources available online, a shift in perspective from a monomodal point of view (e.g., textbook, module) is needed for teachers to create a more meaningful learning atmosphere in the virtual classroom. Further studies on CMC, focusing on multimodality through different channels, using task-based instruction frameworks, are needed with participants coming from wider cultural and linguistic backgrounds. This might help uncover how nonverbal cues (e.g., emojis, gestures, and gaze) are utilized across cultures. Additionally, microanalysis of how students utilize nonverbal cues in tandem with verbal cues in various discourse functions during task interactions needs to be done holistically. Finally, investigating the effect of integrating social media and form-focused tasks into the syllabus by integrating the use of verbal and nonverbal cues is important to connect pedagogical practice to language use in real-world situations.

Acknowledgment

We would like to thank the editors of this special issue and the anonymous reviewers for their constructive feedback on the earlier version of this article. This project was funded by the Ministry of Religious Affairs, Republic of Indonesia in cooperation with *Lembaga Pengelola Dana Pendidikan* (LPDP) under *"Program 5000 Doctor"* and Curtin University, Australia.

References

Aloraini, N. (2018). Investigating Instagram as an EFL learning tool. *Arab World English Journal*, *4*(4), 174–184. https://doi.org/10.24093/awej/call4.13

Aloraini, N., & Cardoso, W. (2020). Social media in language learning: A Mixed-methods investigation of students' perceptions. *Computer Assisted Language Learning*, 1–24. https://doi.org/10.1080/09588221.2020.1830804

Castañeda, D. A. (2019). Improving conversational interactions with task-based activities in a Spanish as a second language class. *Computer Assisted Language Learning*, 34(8), 1154–1181. https://doi.org/10.1080/09588221.2019.1666149

Castrillo de Larreta-Azelain, M. D. (2013). Learners' attitudes toward collaborative writing in e-learning classes: A Twitter project for German as a foreign language. *ReaL Revista Electronica de Linguistica Aplicada*, S127+. https://link.gale.com/apps/doc/A362274086/AONE?u=googlescholar&sid=bookmark-AONE&xid=4582bbd4.

Chang, C. Y.-h. (2016). EFL reviewers' emoticon use in asynchronous computer-mediated peer response. *Computers and Composition, 40*, 1–18. https://doi.org/10.1016/j.compcom.2016.03.008

Chen, J. C. (2018). The Interplay of tasks, strategies and negotiations in Second Life. *Computer Assisted Language Learning*, *31*(8), 960–986. https://doi.org/10.1080/09588221.2018.1466810

Chen, J. C., & Brown, K. L. (2012). The Effects of authentic audience on English as a second language (ESL) writers: A Task-based, computer-mediated approach. *Computer Assisted Language Learning, 25*(5), 435–454. https://doi.org/10.1080/09588221.2011.606224

Côté, S., & Gaffney, C. (2018). The Effect of synchronous computer-mediated communication on beginner L2 learners' foreign language anxiety and participation. *The Language Learning Journal*, 1–12. https://doi.org/10.1080/09571736.2018.1484935

Creswell, J. W. (2012). *Educational research: Planning, conducting, and evaluating quantitative and qualitative research* (4th ed.). Pearson Education.

Creswell, J. W., & Creswell, J. D. (2018). *Research design: Qualitative, quantitative, and mixed methods approaches* (5th ed.). SAGE Publications.

Dewaele, J.-M. (2017). Why the dichotomy 'L1 versus LX user' is better than 'native versus non-native speaker. *Applied Linguistics, 39*(2), 236–240. https://doi.org/10.1093/applin/amw055

Dewaele, J.-M., & MacIntyre, P. D. (2014). The Two faces of Janus? Anxiety and enjoyment in the foreign language classroom. *Studies in Second Language Learning and Teaching, 4*(2), 237–274. https://doi.org/10.14746/ssllt.2014.4.2.5

Dörnyei, Z., & Csizér, K. (2012). How to design and analyze surveys in second language acquisition research. In A. Mackey & S. M. Gass (Eds.), *Research methods in second language acquisition: A Practical guide* (pp. 74–93). Wiley-Blackwell.

Dressman, M. (2020). Multimodality and language learning. In M. Dressman & R. W. Sadler (Eds.), *The Handbook of Informal Language Learning* (pp. 39–55). John Wiley & Sons.

Ellis, R. (2018). *Introducing task-based language teaching*. Curtin University.

Ellis, R., Skehan, P., Li, S., Shintani, N., & Lambert, C. (2019). *Task-based language teaching: Theory and practice*. Cambridge University Press. https://doi.org/10.1017/9781108643689.006

Erarslan, A. (2019). Instagram as an education platform for EFL learners. *TOJET: The Turkish Online Journal of Educational Technology, 18*(3), 54–69. http://www.tojet.net/articles/v18i3/1835.pdf

González-Lloret, M., & Ortega, L. (2014). Toward technology-mediated TBLT: An Introduction. In M. González-Lloret & L. Ortega (Eds.), *Technology-mediated TBLT: Researching Technology and Tasks* (pp. 1–18). John Benjamins Publishing Company.

Guichon, N., & Cohen, C. (2016). Multimodality and CALL. In F. Farr & L. Murray (Eds.), *The Routledge Handbook of Language Learning and Technology* (pp. 509–521). https://doi.org/10.4324/9781315657899

Gurzynski-Weiss, L., & Baralt, M. (2013). Exploring learner perception and use of task-based interactional feedback in FTF and CMC modes. *Studies in Second Language Acquisition, 36*(1), 1–37. https://doi.org/10.1017/s0272263113000363

Hampel, R., & Stickler, U. (2012). The Use of videoconferencing to support multimodal interaction in an online language classroom. *ReCALL, 24*(2), 116–137. https://doi.org/10.1017/s095834401200002x

Huang, X. (2018). Improving communicative competence through synchronous communication in computer-supported collaborative learning environments: A Systematic review. *Education Sciences, 8*(1). https://doi.org/10.3390/educsci8010015

Hung, Y.-W., & Higgins, S. (2015). Learners' use of communication strategies in text-based and video-based synchronous computer-mediated communication environments: Opportunities for language learning. *Computer Assisted Language Learning, 29*(5), 901–924. https://doi.org/10.1080/09588221.2015.1074589

Jewitt, C. (2014). An Introduction to multimodality. In C. Jewitt (Ed.), *The Routledge handbook of multimodal analysis* (2nd ed., pp. 15–30). Routledge.

Jin, L. (2017). Digital affordances on WeChat: Learning Chinese as a second language. *Computer Assisted Language Learning, 31*(1–2), 27–52. https://doi.org/10.1080/09588221.2017.1376687

Kendon, A. (2004). *Gesture: Visible action as utterance*. Cambridge University Press.

Kost, C. R. (2008). Use of communication strategies in a synchronous CMC environment. In S. S. Magnan (Ed.), *Mediating Discourse Online* (pp. 153–190). John Benjamins Publishing Company.

Lee, H., Hampel, R., & Kukulska-Hulme, A. (2019). Gesture in speaking tasks beyond the classroom: An Exploration of the multimodal negotiation of meaning via Skype videoconferencing on mobile devices. *System, 81*, 26–38. https://doi.org/10.1016/j.system.2018.12.013

Lee, L. (2016). Autonomous learning through task-based instruction in fully online language courses. *Language, Learning & Technology, 20*(2), 81–97. https://doi.org/10125/44462

Lee, L., & Markey, A. (2014). A Study of learners' perceptions of online intercultural exchange through Web 2.0 technologies. *ReCALL, 26*(3), 281–297. https://doi.org/10.1017/s0958344014000111

Levak, N., & Son, J. B. (2017). Facilitating second language learners' listening comprehension with Second Life and Skype. *ReCALL, 29*(2), 200–218. https://doi.org/10.1017/S0958344016000215

Li, L., & Yang, Y. (2018). Pragmatic functions of emoji in internet-based communication---A Corpus-based study. *Asian-Pacific Journal of Second and Foreign Language Education, 3*(1), 1–12. https://doi.org/10.1186/s40862-018-0057-z

Lin, H. (2015). A Meta-synthesis of empirical research on the effectiveness of computer-mediated communication (CMC) in SLA. *Lanaguage, Learning & Technology, 19*(2), 85–117.

Long, M. H. (2015). *Second language acquisition and task-based language teaching*. Wiley-Blackwell.

Miles, M. B., Huberman, A. M., & Saldana, J. (2014). *Qualitative data analysis: A Methods sourcebook* (3rd ed.). SAGE Publications.

NapoleonCat. (2021). *Instagram users in Indonesia July 2021*. https://napoleoncat.com/stats/instagram-users-in-indonesia/2021/07/

Negueruela, E., & Lantolf, J. P. (2008). The Dialectics of gesture in the construction of meaning in second language oral narratives. In S. G. McCafferty & G. Stam (Eds.), *Gesture: Second language acquistion and classroom research* (pp. 88–106). Routledge.

Norris, S. (2004). *Analyzing multimodal interaction: A Methodological framework*. Routledge.

Nurhayati-Wolff, H. (2021a). *Penetration of leading social networks in Indonesia as of Q3 2020*. Statista. https://www.statista.com/statistics/284437

Nurhayati-Wolff, H. (2021b). *Share of Instagram users in Indonesia as of April 2021, by age group*. Statista. https://www.statista.com/statistics/1078350

Park, S., & Wu, M. (2020). Instagram as a learning and motivational tool for freshmen English classes at a private Japanese university. *Asian EFL Journal, 24(5)*, 65–107. https://www.elejournals.com/asian-efl-journal/volume-24-issue-5-september-2020/

Prabhu, N. S. (1987). *Second language pedagogy*. Oxford University Press.

Prihantoro, E., & Zulizilah, S. (2017). Instagram as promotional media for Indonesian local brand and business. In W. Strielkowski (Ed.), *Advances in Social Science, Education and Humanities Research* (pp. 195–198). Atlantis Press.

Rassaei, E. (2017). Video chat vs. face-to-face recasts, learners' interpretations and L2 development: A Case of Persian EFL learners. *Computer Assisted Language Learning, 30*(1–2), 133–148. https://doi.org/10.1080/09588221.2016.1275702

Reinhardt, J. (2018). Social media in second and foreign language teaching and learning: Blogs, wikis, and social networking. *Language Teaching, 52*(1), 1–39. https://doi.org/10.1017/s0261444818000356

Samuda, V., & Bygate, M. (2008). *Tasks in second language learning*. Palgrave Macmillan.

Satar, H. M. (2013). Multimodal language learner interactions via desktop videoconferencing within a framework of social presence: Gaze. *ReCALL, 25*(1), 122–142. https://doi.org/10.1017/s0958344012000286

Satar, H. M. (2016). Meaning-making in online language learner interactions via desktop videoconferencing. *ReCALL, 28*(3), 305–325. https://doi.org/10.1017/s0958344016000100

Satar, H. M., & Ozdener, N. (2008). The Effects of synchronous CMC on speaking proficiency and anxiety: Text versus voice chat. *The Modern Language Journal, 92*(iv), 595–613. https://doi.org/10.1111/j.1540-4781.2008.00789.x

Vandergriff, I. (2016). *Second-language discourse in the digital wold: Linguistics and social practices in and beyond the networked classroom*. John Benjamins.

Walsh, S. (2021). *The Top 10 social media sites & platforms 2021*. https://www.searchenginejournal.com/social-media/biggest-social-media-sites/#close

Wigham, C. R., & Satar, H. M. (2021). Multimodal (inter)action analysis of task instructions in language teaching via videoconferencing: A case study. *ReCALL, 33*(3), 195–213. https://doi.org/10.1017/s0958344021000070

Yudhiantara, R. A., & Nuryantini, A. Y. (2019). Instagram-assisted language learning in Islamic higher education: Toward online collaboration. *Journal of English Education and Linguistics Studies (JEELS), 5*(2), 189–210. https://doi.org/10.30762/jeels.v5i2.738

Ziegler, N. (2016). Taking technology to task: Technology-mediated TBLT, performance, and production. *Annual Review of Applied Linguistics, 36*, 136–163. https://doi.org/10.1017/s0267190516000039

Chiang, Y. V., Cheng, Y.-W., & Chen, N.-S. (2023). Improving Language Learning Activity Design through Identifying Learning Difficulties in a Platform Using Educational Robots and IoT-based Tangible Objects. *Educational Technology & Society*, *26*(3), 84-100. https://doi.org/10.30191/ETS.202307_26(3).0007

Improving Language Learning Activity Design through Identifying Learning Difficulties in a Platform Using Educational Robots and IoT-based Tangible Objects

Yueh-hui Vanessa Chiang[1], Ya-Wen Cheng[2] and Nian-Shing Chen[3*]

[1]National Chengchi University, Taiwan // [2]Asia University, Taiwan // [3]National Taiwan Normal University, Taiwan // vanessa.chiang@gmail.com // yawencheng@asia.edu.tw // nianshing@gmail.com

[*]Corresponding author

ABSTRACT: Understanding the obstacles and causes students faced while learning with new technologies is the key to inform effective instructional designs. To achieve this aim, this study conducted a qualitative video analysis on language learners' observable behaviors when they took part in learning activities supported by the technology of robots and IoT-based tangible objects. Insightful findings and instructional implications emerge from the attempt to explore learners' learning process in terms of the obstacles learners encountered and the causes of the obstacles. Based on the findings and implications, eight instructional guidelines are proposed for teachers/instructional designers to design effective language learning activities with robots and IoT-based tangible objects. This study contributes to the literature on enhancing learning and teaching by integrating educational robots and IoT-based tangible objects in the field of robot assisted language learning (RALL).

Keywords: Contextualized multimodal language learning, Robot-assisted language learning, IoT-based tangible objects, Qualitative video analysis, Instructional design guidelines

1. Introduction

The rapid innovation and improvement of information technology expands its application to facilitating learning and teaching in various educational domains. One of the major application areas is English as a foreign language (EFL) learning (Cheng et al., 2018). In traditional EFL learning settings, learners usually face two main challenges to master English. The first one is lacking learning contexts that learners can have sufficient exposure to English in daily life, particularly for advancing their listening skills. The second one is lacking opportunities that learners can apply English to real-life situations for enhancing meaning making in English learning process (Chamot, 2004; DeKeyser, 2005; Seedhouse, 2017). As an endeavor to overcome such challenges, this study was grounded in an EFL learning environment integrating humanoid robots and Internet of Things (IoT) based tangible objects as instructional technologies to enhance immersion and engagement for EFL learners at elementary school level. IoT technologies refer to real-world tangible objects, which are embedded with digital sensors for collecting relevant data and connected with other objects via network technologies like intranet or internet for storing, exchanging, and analyzing data to fulfill specific purposes (Madakam et al., 2015).

Humanoid robot can be used to foster active learning (Kim & Baylor, 2006) by playing a role in language learning activities, socially interacting with learners via its utterances, facial expressions and gestures to facilitate learners' learning process. In this study, such social interaction with robots intended to engage young language learners to repetitively practice English listening skills. The inclusion of tangible objects embedded with IoT sensors can serve as embodied learning materials related to learning contexts to increase the degree of modalities that supports situated learning (Hung et al., 2015; Pasfield-Neofitou et al., 2015). In a language learning environment supported by technology without embodiment, learners may just sit on chairs in front of computer monitors to interact with two-dimensional digital learning materials. This study incorporating tangible objects supported by IoT technologies into the learning environment attempted to expand two-dimensional digital learning materials to three-dimensional cyber-physical mixed learning contents. Such learning environment affords language learners to interact with tangible learning materials via various body movements in learning process, which supports embodied cognition that has been regarded as beneficial for helping learning retention and meaning making in learning contexts requiring real-life application scenarios (Shilling, 2017). From the perspective of cognitive science, learning involves the process of cognition. The viewpoint of embodied cognition emphasizes the mutual influences between brain and body in the cognitive process. Learning activities using technologies to elicit learners' body movements can facilitate learning in terms of enriching mental representations and enhancing retention of learned information (Chao et al., 2013). Thus, this study intended to establish a contextualized multimodal language learning environment supporting both social interaction and embodied cognition with a novel integration of robots and IoT-based tangible objects.

Teachers are experts in the subjects they teach. Most teachers are also good at using technology with which they are already familiar to deliver instruction. However, when it comes to integrating new technology into learning and teaching, educators have noticed a challenge that not every teacher is always ready to incorporate state-of-the-art instructional technology into learning effectively and efficiently (Becker, 1994; Cuban, 2001; Hadley & Sheingold, 1993; Tondeur et al., 2013). Ertmer and Ottenbreit-Leftwich (2013) suggested that the integration of educational technology should pay attention to the instructional aspect to facilitate learning and teaching for students and teachers, rather than technological deployment merely. Thus, supports for helping teachers integrate new technology to fulfil their instructional goals are needed (Batane & Ngwako, 2017). A set of instructional guidelines may serve the purpose. Moreover, Wells (2001) suggested that learning process making up the learning experience is likely to provide further insights for enhancing learning. As such, the objective of this study is to provide instructional guidelines based on analyzing students' observable behaviors when they participated in the language learning activities supported by the learning environment with robots and IoT-based tangible objects.

The novel integration of emerging technologies may arise new instructional concerns to be discovered. Video recording students' learning process makes the attempt of gaining insights from the course of learning feasible (Janík et al., 2009). Identifying the obstacles students encountered initiates our video analyses from where students got stuck in their learning process. Moreover, exploring the causes of the obstacles can inform teachers how to support students' language learning in such learning context. Thus, this study aims at addressing the following three research questions: (1) what obstacles did language learners encounter when participating in learning activities with robots and IoT-based tangible objects? (2) what causes led to the obstacles? (3) what can the identified obstacles and causes inform the design of language learning activities with robots and IoT-based tangible objects?

2. Literature review

2.1. Recent relevant studies on educational robots and IoT-based tangible objects

As emerging technologies, robots and IoT have drawn different levels of attention from researchers in education. Prior studies on robots in education are relatively abundant and considered favorable for supporting young learners' language learning in terms of vocabulary learning, language production, listening skills, and oral interaction, as well as reducing learning anxiety (Alemi et al., 2014; Alemi et al., 2015; Cheng et al., 2021; Cheng et al., 2018; Hung et al., 2015; Lin et al., 2022; Pachidis et al., 2019). Prior studies on using tangible objects with IoT technologies in education are relatively few (Domínguez & Ochoa, 2017; Seedhouse, 2017; Skulmowski et al., 2016; Xie et al., 2008). Noteworthy, Seedhouse and his colleagues (2017) conducted a series of studies using IoT-based tangible objects in kitchen with a task-based pedagogical approach to support foreign language learning and reported such environment can be used to facilitate listening skills and vocabulary learning (Pallotti et al., 2017; Roos et al., 2017). Even though the application of IoT technologies is not as popular as educational robots yet, researchers start to recognize its potential for constructing a personalized smart learning environment (Khaddage et al., 2016; Gul et al., 2017; Ramlowat & Pattanayak, 2019). Given that current applications of educational robots and IoT-based tangible objects as learning technologies are on different stages, few studies integrated these two technologies into the same learning contexts (Tanaka et al., 2007). Our study integrating robots and IoT tangible objects served as a pioneering attempt to establish a contextualized multimodal language learning environment leveraging the advantages of robots, IoT, and tangible objects.

Among recent works of using robots and/or IoT objects in educational contexts, many of them supported that using these technologies can attract learners' attention and their presence is acceptable to younger learners (Chen et al., 2011; Hyun et al., 2010; Kanda et al., 2004; Kanda et al., 2007; Kozima et al., 2009; Leite et al., 2009; Leite et al., 2013 Tanaka et al., 2007). However, most of the prior studies tended to conduct experimental inquiries and regarded the use of educational robots and/or IoT objects as treatments in research design to examine the effects of the technologies on learning outcome via summative measurements at a certain time point. Few of them investigated the use of educational robots and/or IoT objects as learning technologies from the formative aspect throughout the process of learning activities (Benitti, 2012; Karim et al., 2015; Li, 2015; Liu et al., 2016; Mubin et al., 2013; Toh et al., 2016; Spolaôr & Benitti, 2017). Therefore, this study attempted to bridge the gap by focusing on analyzing language learners' observable behaviors throughout their learning process.

2.2. The importance of supporting teachers integrating new technology into learning and teaching

Applying emerging technology to learning and teaching may create more opportunities of educational innovation and improvement. However, as Ertmer and Ottenbreit-Leftwich (2013) suggested that the integration of educational technology should pay particular attention to the instructional aspect of how teachers can facilitate learning and teaching with such technology, instead of merely to the development of technology. Simply using new technology in learning activities does not necessarily enhance learning. Jonassen (1996) stated teachers should focus on designing learning activities incorporating technology in facilitating students' learning process. By doing so, the benefits of using technology can be extended to carry out effective and meaningful learning. In order to address this concern, educational researchers proposed that support for teachers to enable their adoption of new technology in learning process is needed especially when more technology is developed and promising for being used in learning contexts (Batane & Ngwako, 2017; Cheng et al., 2021; Ertmer & Ottenbreit-Leftwich, 2013). Additionally, Howland et al. (2012) emphasized helping teachers adopt new teaching practices afforded by new technology should focus on the change of learning process.

This study introducing a novel integration of emerging technology, educational robots and IoT-based tangible objects, may raise new instructional concerns when teachers would like to design effective learning activities with such technology. Responding to suggestions from existing literature, this study attempted to provides guidelines for teachers to design language learning activities with educational robots and IoT-based tangible objects for smoothing the process of adopting new educational technology.

In order to achieve this objective, the guidelines particularly derive from analyzing language learners' observable behaviors when they participated in learning activities with educational robots and IoT-based tangible objects. Analyzing learners' behaviors to identify what may hinder or assist learning process can provide insights for how to engage learners in learning by avoiding hindrance or rendering assistance (Vygotsky & Cole, 1978; Wells, 2001). Accordingly, the analyses started from identifying the obstacles that learners encountered when they participated in the language learning activities involving educational robots and IoT-based tangible objects. Subsequently, identifying the causes of the encountered obstacles may inspire what needs to be taken into consideration when teachers want to redress the obstacles.

3. The design of learning activities

3.1. The robot and IoT-based tangible objects language learning environment

This study was conducted in a technology-enhanced language learning environment that integrates robots, tablets, and tangible objects with IoT sensors and is driven by scripts composed by instructional designers with a visual programming tool. The learning environment was constituted by the frontend and backend system components. The frontend system components refer to the devices, such as robots, tablets, and tangible objects, with which learners actually interact. The backend component called script editor is a visual programming platform on which teachers/instructional designers work for designing learning activity. How the frontend elements interact with learners depends on the scripts composed with the backend script editor.

The script editor is a visual programming website adapted from Google Blocky (see Figure 1). Teachers/instructional designers compose scripts with this visual programming tool to design learning activities involving the frontend system components and to manage the learning materials used in the learning activities. The scripts created by teachers/instructional designers are saved on a cloud server beforehand and deployed to the frontend components to steer what and how they perform when the corresponding learning activities take place.

The frontend system components include a robot, a tablet, and a set of tangible objects embedded with IoT sensors as shown in Figure 2. The role the robot assumes may vary according to teachers'/instructional designers' decisions to carry out their instructional design. Also, an additional tablet, dedicated to displaying multimedia materials is accompanied with the robot to increase the richness of information presented to learners. The IoT-based tangible objects used for this study is a set of supermarket toys, including a supermarket shelf with many tangible food or stationery models (see Figure 2). In addition to the multimedia materials displayed on the tablet, the tangible toys are three-dimensional learning materials that support embodiment for learners in this multimodal learning environment.

Figure 1. The backend script editor

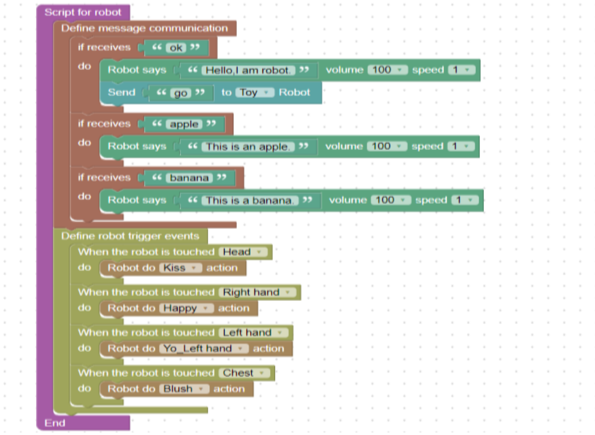


Figure 2. The frontend system components

Figure 3. A toy object with a NFC tag attached

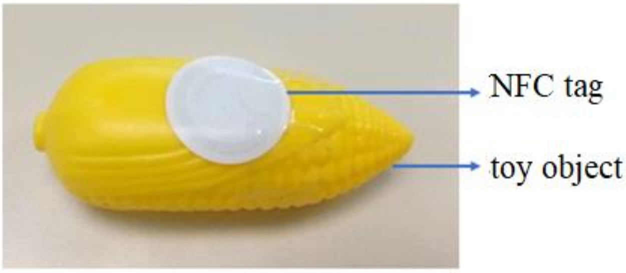
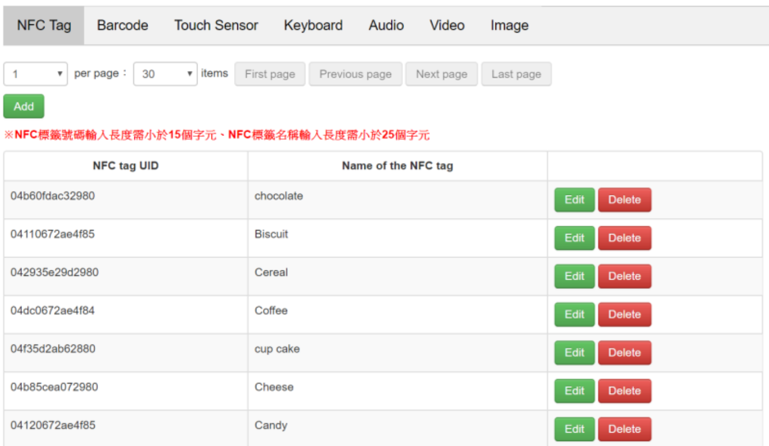


Figure 4. Information of NFC tags and associated toys



Content management

NFC Tag | Barcode | Touch Sensor | Keyboard | Audio | Video | Image

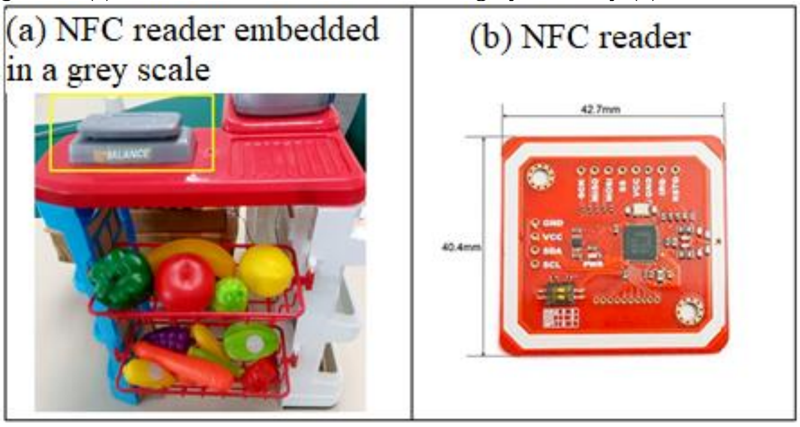
1 per page： 30 items | First page | Previous page | Next page | Last page

Add

※NFC標籤號碼輸入長度需小於15個字元、NFC標籤名稱輸入長度需小於25個字元

NFC tag UID	Name of the NFC tag	
04b60fdac32980	chocolate	Edit Delete
04110672ae4f85	Biscuit	Edit Delete
042935e29d2980	Cereal	Edit Delete
04dc0672ae4f84	Coffee	Edit Delete
04f35d2ab62880	cup cake	Edit Delete
04b85cea072980	Cheese	Edit Delete
04120672ae4f85	Candy	Edit Delete

Figure 5. (a) An NFC reader embedded into a grey scale toy (b) An NFC reader

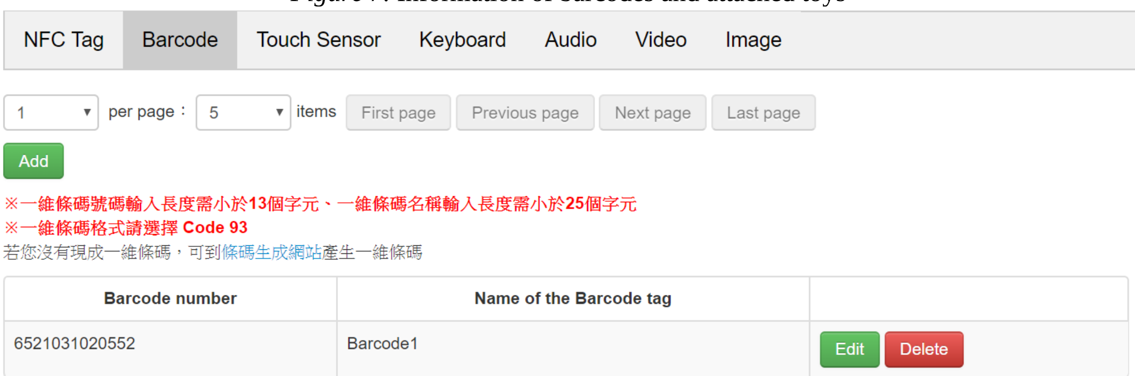


The supermarket toys are adapted to accommodate two types of sensors. First, Near Field Communication (NFC) tags and readers: each toy object is attached with an NFC tag (see Figure 3). The corresponding information of which NFC tag is associated with which tangible toy object is set and kept in the backend script editor (see Figure 4). An NFC reader is embedded into a grey scale toy as shown in Figure 5(a)(b). When learners place a toy object on the grey scale, the robot can tell which object it is and trigger corresponding actions.

Second, one-dimension barcodes and the reader: similar to NFC tags, one-dimension barcodes are stuck on some tangible objects (see Figure 6(a)). Learners can hold a barcode reader to scan the barcodes (see Figure 6(b)). A corresponding table of which barcode represents which toy object is also set in script editor (see Figure 7).

Figure 6. (a) A barcode stuck on a toy object (b) A barcode reader

Figure 7. Information of barcodes and attached toys



NFC Tag | Barcode | Touch Sensor | Keyboard | Audio | Video | Image

1 per page： 5 items First page Previous page Next page Last page

Add

※一維條碼號碼輸入長度需小於13個字元、一維條碼名稱輸入長度需小於25個字元
※一維條碼格式請選擇 Code 93
若您沒有現成一維條碼，可到條碼生成網站產生一維條碼

Barcode number	Name of the Barcode tag	
6521031020552	Barcode1	Edit Delete

3.2. The learning activities

This study followed the framework suggested by Hubbard (1992) in terms of integrating technology into language learning to design task-based activities, which has been considered suitable for enhance language learning (Pica et al., 2006; Morales, 2017). The language learning activities in this study consisted of four sessions in four weeks. Each session included several tasks for learners to complete one by one. The learning objective of this series of learning activities was to improve learners' English listening skills. Responding to Hubbard's (2017) suggestion of using emerging technologies to establish a three-dimensional immersive environment with digital sensors for learning language listening, this study used robots and IoT technologies for this purpose. With such foundation, most of the instruction and hints of the tasks were intendedly presented by the robot in the audio format, supplemented with some visual materials on the tablet. The vocabularies and conversations used in the learning activities were designed in alignment with the content of the fourth graders' English textbooks in Taiwan and confirmed by the participants' English teachers before the learning activities took place.

The goal for the learning activity in the first week was to prepare learners for the subsequent activities by helping them get familiar with how to use the two types of sensors (i.e., the NFC and the barcode). The goal for the week 2 learning activity was to help learners practice their listening skills of vegetables, fruits, and food. The learning activity of the third week attended to cover the vocabularies of stationery, colors, clothes and drinks in addition to fruits. The goal of week 4 learning activity was to help learners practice their listening skills of colors. Several tangible objects embedded with sensors were used in the learning activities as learning materials, representing the learning content. Such learning activities were not designed to replace the existing English class sessions. Instead, before the learning activities took place, the participants' English teachers previewed the learning activities and acknowledged the potential for supplementing the existing English classes, enhancing students' learning motivation, and further improving students' English listening skills. In addition, the teachers suggested to incorporate more learning supports, such as oral encouragement and visual hints assisting learners in connecting the vocabularies and the corresponding objects, into the learning activities. Teachers' suggestions were all implemented in the learning activities.

Figure 8. (a) The robot provides instruction of a task (b) The learner places the found object on grey scale (c) The learner scans the found object with barcode reader

(a) (b)

(c)

The general process occurring in the four learning activities was that the robot started to provide instruction describing the target object to complete a task to learners as shown in Figure 8(a). Then, the learners worked on finding the target object based upon their understanding of the instruction. Learners would present the tangible object they found by either placing it on the grey scale or scanning it with the barcode reader as shown in Figure 8(b) and Figure 8(c). The robot would determine if the presented object is the target one and provide corresponding feedback to the learners to indicate their success or failure. In the case of success, the robot would proceed to give instruction of the next task until all tasks were completed. In the case of failure, the robot would encourage the learners to try again. Upon the completion of all tasks, the robot would render a final congratulation to the learners.

4. Method

From the sociocultural perspective (Vygotsky & Cole, 1978), this study attempted to analyze students' observable behaviors during task performance in the real learning situations supported by the learning environment with robots and IoT-based tangible objects to address the three research questions with an inductive qualitative approach (Lincoln & Guba, 1985).

4.1. Participants

This study took place in an elementary school in central Taiwan. Ten fourth graders participated in this study with their parents'/legal representatives' consent. Participants' age ranged between 9 and 10 years old. Five of them were female; five were male. As shown in Figure 9, a classroom was allotted to be the venue of this study. Sound-proofing panels were used to set up 10 cubicles. Participants mainly worked within their assigned cubicles during the learning activities. In each cubicle, a robot with a tablet and a set of supermarket toys embedded with IoT sensors and readers were equipped. Participants attended one learning session, lasting 20-30 minutes, per week for four consecutive weeks in 2019.

Figure 9. Research site

4.2. Data sources and analysis

The primary data source of this study was the video recordings of participants' learning process. To collect the primary data source, each cubicle was equipped with a camera to record what happened in the cubicle during the learning session. The observational field notes served as a secondary data source.

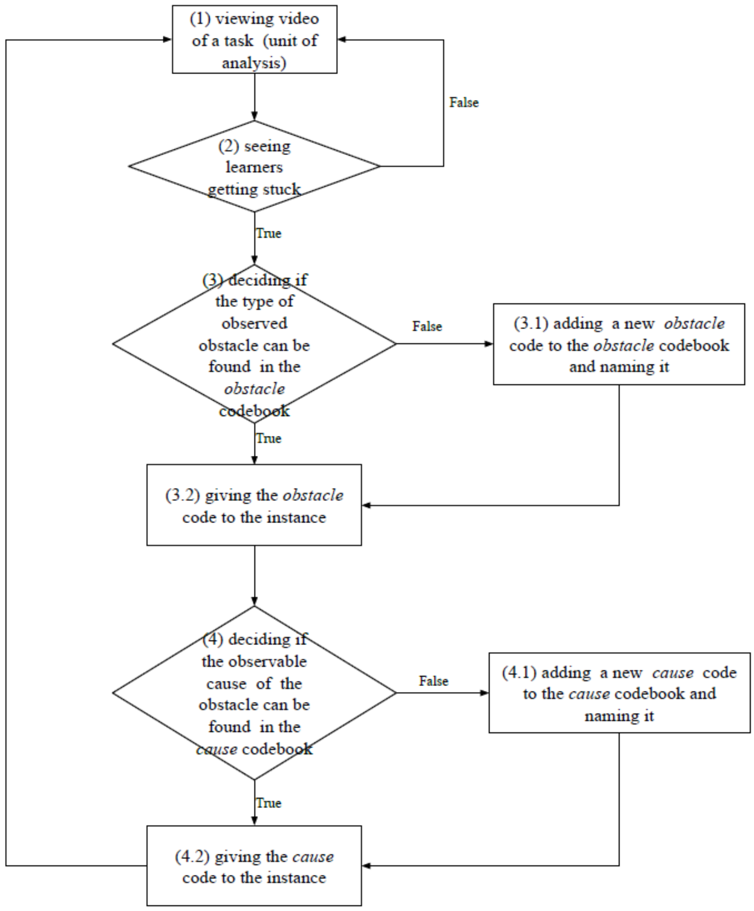
The technique of video analysis was used to analyze the data. All video clips were viewed for the first round to select valid clips to proceed to the next round of analysis. The criteria for selecting valid clips included: (a) the visibility of the participants in the video clips, (b) the audio quality of the video clips, and (c) the visibility of the objects used in the learning activity. The units of analysis for the subsequent analysis were the tasks that the participants were assigned to complete in every learning activity. A total of 27 valid video clips were selected and 135 tasks were analyzed to identify obstacles and causes to address the first two research questions.

The open coding technique for qualitative inquiries (Strauss & Corbin, 1990) was adopted to generate the codebooks in terms of *obstacles* and *causes*. Figure 10 presents the procedure of identifying *obstacles* and *causes* from video recordings in the following steps:

- Step (1): the researchers viewed the video clip of a task.
- Step (2): when seeing the participants getting stuck, the researchers regarded such instance as an obstacle and proceeded to the next step, otherwise, continuing to viewing other video clips.
- Step (3): deciding if the observed obstacle belonged to an existing type of obstacle in the *obstacle* codebook. If not, the researchers proceeded to step (3.1), otherwise, proceeding to step (3.2).
    - Step (3.1): adding a new *obstacle* code which can fit the new instance to the *obstacle* codebook, labelling the new code, and continuing to step (3.2).
    - Step (3.2): giving the *obstacle* code to the instance.
- Step (4): deciding if the observable cause of the obstacle belonged to an existing type of cause in the *cause* codebook. If not, the researchers proceeded to step (4.1), otherwise, proceeding to step (4.2).
    - Step (4.1): adding a new *cause* code which can fit the new instance to the *cause* codebook, labelling the new code, and continuing to step (4.2).
    - Step (4.2): giving the *cause* code to the instance.

Then, the researchers repeated the step (1) to (4) for all selected video clips. As data analysis proceeded, a set of codes of *obstacles* and *causes* emerged and were continuously compared with new instances to see if modifications were needed. Two researchers followed the procedure to code the data separately first and then reach their consensus in terms of their coding results. The Cohen's kappa coefficient of *obstacle* coding was 0.929 and the coefficient of *cause* coding was 0.873.

Figure 10. The procedure of identifying obstacles and causes from video clips



Before answering the research questions, the researchers confirmed the effectiveness of the learning activities with robots and IoT-based tangible objects on learners' learning outcome measured by pre-test and post-test of learners' English listening skills. A Shapiro-Wilk test of normality indicated pre-test (p = .000) and post-test (p = .008) scores were not normally distributed. Consequently, a Wilcoxon Signed-Ranks test indicated the median of post-test scores, 82.5, was significantly higher than the median of pre-test scores, 80 (Z = -2.000, p = .046 < .05). We, therefore, concluded there was a significant positive effect of the learning activities with robots and IoT-tangible objects on participants' English listening test scores. With confirming the effectiveness on learning outcome, the detailed analysis of learners' learning process aimed for improving the instructional design, making learning experience even better.

5. Results

Among the 135 tasks, 84 instances were coded with obstacles during participants' learning process. The results addressing the first two research questions are reported in this section. Five categories of obstacles that prevented learners from completing the tasks and 7 categories of causes leading to the obstacles emerged from data analysis. Table 1 summarizes the identified categories of obstacles and causes along with definitions and examples. The first and second columns show the names of obstacle categories and the features that were observed in the video recordings of the corresponding obstacles. The third and fourth columns present the observable causes of the corresponding obstacles and their features seen in the video clips. The fifth column provides the examples of the obstacles and causes extracted from the data. The naming convention of referring to a specific raw data of video follows the format of [S##-W#-T#]; where S## indicates the participant ID, W# indicates the number of the week, and T# presents the number of the task in each week's learning activity. For example, [S05-W4-T2] indicates the case of participant S05's second task in week 4.

The first four categories were ordered by the sequence of which they could happen in a given task. At the beginning of each task, when the task instruction was provided to learners mainly by the robot in audio mode, the obstacles of failing to understand the assigned task (O1) and missing the instruction of task (O2) and their causes could occur. After the instructions of tasks were given, participants started to find the target objects according to their understanding of the instructions and the obstacle of failing to find the target toy object (O3) could happen. After participants found the objects to complete the tasks, the obstacle of failing to present the found toy objects (O4) could occur. As to the fifth obstacle of being interfered by contextual factors (O5), it could happen throughout the whole process of completing the tasks.

Table 1. The identified obstacles and causes

Obstacles	Definition of obstacles	Causes	Definition of causes	Examples
O1. Fail to understand the assigned task	(a) The participants were present in the cubicle as the robot rendered the instruction of the assigned task, and (b) the participants seemed not completely understand what the assigned task was.	C1. Learners' English ability	Even though the participants seemed to listen to the instruction, their subsequent behaviors observed in the video recordings indicated that they did not exactly comprehend the instruction and did not know for which target object they were looking.	Participant S05 worked on the task of finding and scanning the orange fingerprint with the barcode reader. After the robot provided the instruction of collecting an orange fingerprint, he tried to place the fruit orange on the grey scale twice and failed to complete the task. Then, he tried many different toy objects until another participant came to his cubicle pointing out where the orange fingerprint was [S05-W4-T2].
		C2. Misconceiving the instruction	The participants seemed to misconceive the visual/audio feedback or hints in the learning activities.	After participant S07 completed the task of placing the starfruit on the grey scale, audio and visual feedback both indicated her completion of the task. The visual feedback left a picture of starfruit on the tablet screen. Then the robot started to provide the audio instruction of the next task, placing a pencil on the grey scale. While participant S07

				seemed to listen to the instruction, she looked at the tablet screen as well. Her first attempt was to place the starfruit on the grey scale again and failed to complete the task [S07-W3-T2].
O2. Miss the instruction of task	The participants missed the instruction of task so that they did not know what the assigned task was.	C3. Not attended to the instruction	(a) The participants were not physically present in the cubicle as the robot rendered the instruction of the assigned task. (b) The participants were physically present in their cubicles, but they were seen that their attention was distracted by other things.	(a) Participant S08 was absent from her cubicle when the robot was giving instruction. After she returned to her cubicle, she looked puzzled at what was going on until one researcher came to give her some hints on the target object of starfruit [S08-W3-T1]. (b) Participant S03 looked excited after he completed the first task. As the robot was giving the instruction of the second task, he was still immersed in the excitement of collecting the green fingerprint by holding and looking at the green pepper with the green fingerprint and did not pay attention to the new instruction [S03-W4-T2].
O3. Fail to find the target toy object	When the participants seemed to know what they were looking for but cannot find the target among the toys	C4. Unable to see the target object	The target object was placed in a location which the participants cannot reach.	Participant S10 attempted to find the orange fingerprint to complete a task. However, the orange fingerprint appeared on the highest shelf and was covered by another toy object. Participant S10 was not tall enough to reach the highest shelf. She kept looking for the orange fingerprint until another participant who was taller than participant S10 help remove the object covering the orange fingerprint [S10-W4-T2].
		C5. Unable to recognize the target object	The participants failed to identify the right toy object representing the English vocabulary of target object.	Participant S07 worked on placing the green apple on the grey scale to complete a task. She tried many times with green pepper without recognizing that the toy object she held was not green apple [S07-W2-T3].
O4. Fail to present the found toy object	Participants had found the target object to complete the assigned task but failed to present it via input devices	C6. Flawed operation of input devices	(a) The participant presented the object with wrong input device. (b) Ineffective operation of the	(a) Participant S08 would like to complete the task of scanning the barcode on a jar of milk. However, instead of using the barcode reader to scan the barcode on the milk

			input device	jar, she placed the jar of milk on the grey scale [S08-W3-T5]. (b) The participants used the right input device to present the found object; however, their operations of the input devices failed to take effect. They needed to repeat the operation to complete their tasks.
O5. Interfered by contextual factors	Even though participants stayed with the robot as it was giving instruction of the assigned task, some contextual issues prevented them from proceeding with the task	C7. Surrounding noises	Two sources of noises from the surroundings: (a) the noise from the school-wide environment, and (b) the noise from the research context, i.e., the participants	(a) As participant S10 was listening to the task instruction, a school teacher made an announcement via broadcasting system. The volume of the broadcast was too loud for participant S10 to hear the instruction clearly. In the video recording, she was observed to incline forward to the robot to hear the instruction. After the broadcast, the robot had finished the instruction [S10-W2-T5]. (b) The cubicle in the research site was not a completely closed area. If the participants talked too loudly, their voices can be heard by all participants in the classroom.

Table 2. The counts of identified obstacles and causes

Obstacles	Causes	W1	W2	W3	W4	Sub total	Total
O1. Fail to understand the assigned task	C1. Learners' English ability	1	8	19	6	34	41
	C2. Misconceiving the instruction	0	0	3	4	7	
O2. Miss the instruction of task	C3. Not attended to the instruction	1	2	6	4	13	13
O3. Fail to find the target toy object	C4. Unable to see the target object	1	0	0	18	19	21
	C5. Unable to recognize the target object	1	1	0	0	2	
O4. Fail to present the found toy object	C6. Flawed operation of input devices	2	1	3	0	6	6
O5. Interfered by contextual factors	C7. Surrounding noises	0	2	1	0	3	3
Total		6	14	32	32	-	84

Table 2 reports the numbers of instances per week of each category. Among the instances coded with obstacle of failing to understand the assigned task (O1), they were mainly caused by learners' English ability (C1) (34 out of 41). The highest occurrences of obstacles O1 caused by C1 was in the third week (i.e., 19). This finding suggested that the learning content covered in week 3 posed the most challenges to the participants' existing English ability. The target words to complete the tasks in week 3 were from the most various categories, including fruits, stationery, clothes, drinks and colors. The participants were less likely to rely on the words completing the preceding tasks as contextual clues to infer the target words of the subsequent tasks. According to Vygotsky's (1962) notion of zone of proximal development, we regarded these challenges as learning opportunities for language learners to expand their English ability. Being seen in the data revealing learners' task performance, the participants eventually overcame the challenges with supports provided by the learning environment. The findings provided us with insights in terms of how to assist learners to overcome challenges in their language learning activities so as to expand their English ability, leading to the instructional guidelines proposed in the next section.

The second highest frequency of obstacles was failing to find the target toy object (O3) because learners cannot see the object (C4) in week 4 (i.e., 18). By reviewing these instances, they were related to the multimodal features provided by the learning environment. As the example of this category shown in Table 1, the learning activity in week 4 asked participants to collect five fingerprints. In the research context, the fingerprint stickers were so thin as tiny pieces of paper that they can be hidden in the places where some participants were unable to reach like the orange fingerprint in the case of [S10-W4-T2]. Also, the 7 instances of obstacle of failing to understand the assigned task (O1) due to learners' misconception of the instruction (C2) shed light on the similar issue of multimodality. As the case of [S07-W3-T2] listed in Table 1 exemplifying this category, when providing feedback to indicate learners' success of completing the task, the instructional designer added visual feedback along with the audio one. In this case, the participant S07 seemed to misperceive the visual feedback of the task she just completed as the instruction of the next task. Although listening to the audio instruction of the next task, she misconceived the picture left from the feedback of the previous task as part of the instruction of the new task, which prevented her from fully comprehending the instruction. These instances informed us in designing multimodal English learning activities, and the guidelines will be presented in the next section.

As to the obstacle of missing the instruction of task (O2) because learners did not attend to the instruction (C3), the instances of this category happened in all 4 weeks. It suggested young language learners may be easily distracted when they participated in learning activities. In this study, the robots provided a function of repeating the instruction of tasks to learners. This function was triggered by touching robots' belly and intended to help learners overcome this obstacle whenever needed. Another way to address this obstacle would be allowing learners to control the flow of the activities. When learners were ready to proceed to the next task, they were in charge of letting the robot know that they would like to listen to the instruction of the next task. The instances of the obstacle of failing to find the target toy object (O3) because learners were unable to recognize the object (C5) occurred only in the first two weeks. This suggested that learners may need some time to get familiar with the tangible objects and associate them with the corresponding English vocabularies. The cases coded with the obstacle of failing to present the found toy object (O4) due to flawed operation of input devices happened in the first three weeks and none in the fourth week. The learning activities of the first three weeks required learners to use either the NFC reader or barcode reader as input devices. However, for the learning activities in the fourth week only required learners to use the barcode reader as the input device. This finding suggested when an IoT-based learning environment allowed learners to use more than one input devices for interaction, learners may need assistance in using the right input device effectively. As for the obstacle of being interfered by surrounding noises (C7 of O5), this can interrupt English learners' learning process, especially when the learning objective was to practice listening ability. This issue may be resolved by equipping headset microphones as sound input/output devices to learners. In addition, the aforementioned functionality of allowing learners to ask the robots to repeat the task instruction when needed can serve as another solution.

6. Instructional guidelines for designing learning activities using robot and IoT-based tangible objects

To address the third research question, the findings of the first two research questions were synthesized to derive instructional guidelines for teachers/instructional designers who may consider designing language learning activities involving robots and IoT-based tangible objects.

During learning process, some challenges with appropriate scaffoldings may be beneficial to expand learners' understanding of the subject matters (Vygotsky, 1962; Wood et al.,1976; Nikolaevskaya, 2017). Hence, the obstacles of failing to understand the assigned task caused by learners' English ability (C1 of O1) may particularly indicate the opportunities of advancing learners' learning and inform the instructional designers how to facilitate learners' learning process by providing scaffoldings when applicable. In order to make this occurs, four guidelines were proposed as follows:

Guideline 1: Teachers/instructional designers should use robots to confirm learners' understanding of the assigned task in the early stages of language learning activities and making applicable rectification in time to help learners stay on the right track, avoiding detours to ineffective learning.

Guideline 2: When tangible objects are used to represent the new learning content, teachers/instructional designers should use robots to strengthen learners' comprehension of learning content by providing assistance in associating new content with the corresponding tangible objects in the language learning context.

Guideline 3: Teachers/instructional designers should avoid confounding new learning content with multiple meanings and use lexical items that are suitable for learners' age in terms of word frequency and familiarity.

Guideline 4: Teachers/instructional designers should use robots to provide functionality that allows learners to control the progress of the learning activities, such as learners can ask robots to repeat task instruction whenever they need, and learners can ask robots to proceed to the subsequent tasks whenever they are ready.

Speaking of facilitating learning, designing some sort of functionality enabling the learning environment to keep track of learners' learning process can be beneficial to trigger applicable assistance. For example, with a function to monitor the count of failure to complete the tasks, when it passes a certain amount, robots can automatically provide more guidance to learners. Moreover, with a function to keep track of time span between learners' attempts to complete the tasks, if it is longer than a certain period of time, robots could start to provide assistance as well. These scaffoldings are appealing for facilitating learning process and constructing an adaptive learning environment that can offer multiple assistance in accordance with learners' learning states, leading to the next guideline.

Guideline 5: Teachers/instructional designers should make robots' behaviors responsive to learners' learning states and offer multiple types of assistance available to language learners.

In addition, the obstacle of failing to present the found toy object due to the flawed operation of input devices (C6 of O4) was mainly related to the familiarity of using new technologies. It had a little to do with learners' English ability but can definitely hinder learners' learning process. Whenever introducing new technologies to learning and teaching, teachers/instructional designers should prepare learners for necessary operations of the new technologies in advance; otherwise, it is possible to embrace emerging technology at the cost of learners' learning effectiveness. In this study, the instructional designer had designed the learning activity in week 1 to help learners get familiar with the new input devices. However, some cases of this category of obstacles still occurred. More cases would be anticipated if not implementing this activity. Thus, the sixth guideline was provided:

Guideline 6: Teachers/instructional designers should make sure that the pre-task activities preparing learners' operation of newly adopted technologies and familiarity with related vocabularies are well designed and thoroughly carried out so that learners can attend to the language learning tasks.

Moreover, the obstacle of failing to find the target toy object caused by being unable to see the target object (C4 of O3) occurred the most in week 4 is noteworthy. A multimodal language learning environment affording textual, audio, spatial, and visual modes of interaction is promising for enabling immersive language learning (Moreno & Mayer, 2007; Kirsh, 2013). Meanwhile, when designing learning activities, as suggested by Skulmowski and colleagues (2016), the instructional designers should consider the appropriateness of multimodal interaction between learners and the learning environment, rather than only thinking about the traditional mono-modal communication. The examples of this obstacle highlight the importance of such consideration. Placing the target objects at a physical location where learners cannot reach makes the objects beyond the scope of learners' normal vision and hinders learners from completing the learning activities. In this regard, neglecting the aspect of three-dimensional spatial modality can even create more obstacles for learning, particular for young children. To avoid this pitfall, a comprehensive consideration including learners' profiles and the features of multiple modalities when designing language learning activities is recommended for instructional designers as below:

Guideline 7: Teachers/instructional designers should take learners' physical attributes into consideration when arranging tangible objects physically.

Furthermore, the obstacle of failing to understand the assigned task caused by misconceiving the instruction (C2 of O1) also relates to multimodality. It sheds light on the concerns about proper sequence of the multimodal materials used in language learning activities and the consistency of messages delivered to learners via multiple modalities. In a language learning environment supporting multimodal materials, confusing arrangement of the materials and inconsistent messages delivered by them at the same time can hold back the advantages of multimodality and even deteriorate the learning process, which results in the eighth guideline:

Guideline 8: When using robots to deliver messages serving the same purpose via different modalities, such as robots' utterances, facial expressions, and gestures, teachers/instructional designers should make sure that these messages are delivered consistently. Likewise, robots should deliver the messages serving different purposes in a way distinct enough for learners to differentiate their references correctly.

7. Conclusion

This study proposed instructional guidelines for teachers/instructional designers to design learning activities involving robots and IoT-based tangible objects for language learning. The guidelines emerged from analyzing learners' learning process to understand what obstacles learners encountered and the causes of the obstacles as learners participated in language learning activities. These attempts may contribute to supporting teachers/instructional designers who are interested in integrating robots and IoT-based tangible objects to design effective language learning activities for learners. By adopting educational robots and IoT-based objects together, this study introduced an innovative English learning environment that can accommodate multimodal pedagogical practices. Such endeavors may open up new opportunities to enhance language learners' exposure to English and to connect what they learned to real-life scenarios.

Even though the qualitative video analysis granted us rich details of learners' learning process, this study has limitation due to the data collection of video recordings. The researchers set up the recording equipment at a location that can cover students' cubicles mostly. However, in the real learning activities, students might walk outside the camera scene. This methodological limitation has been commonly seen in the studies using video recordings as data sources (Janík et al., 2009). To recognize this limitation, the researchers made the findings and suggestions based upon the observable evidences in the video recordings. Another limitation is that this study only included the language learning activities with single learning strategy, task-based learning, in the learning environment with robots and IoT-based toys. The future direction will be conducting more studies with various learning strategies supported by this learning environment to develop more guidelines for teachers/instructional designers. Making the language learning environment using robots and IoT-based tangible objects more applicable and effective to support learners' learning process is the ultimate goal. To make this happen, the researchers recognize the importance of teachers' perspective on using robots and IoT-based tangible objects to design and carry out learning activities. Thus, in addition to proposing a set of instructional guidelines for teachers, the researchers will conduct future studies to understand teachers' perceptions and collect their feedback in terms of using robots and IoT-based tangible objects to enhance students' learning.

Acknowledgement

This research is supported in part by the Ministry of Science and Technology, Taiwan under project numbers MOST 111-2410-H-003 -028 -MY3, MOST-109-2511-H-003-053-MY3, MOST 109-2511-H-004-004, and MOST-108-2511-H-003-061-MY3. This work was financially supported by the "Institute for Research Excellence in Learning Sciences" of National Taiwan Normal University (NTNU) from The Featured Areas Research Center Program within the framework of the Higher Education Sprout Project by the Ministry of Education (MOE) in Taiwan

References

Alemi, M., Meghdari, A., & Ghazisaedy, M. (2014). Employing humanoid robots for teaching English language in Iranian junior high-schools. *International Journal of Humanoid Robotics*, *11*(03), 1450022. https://doi.org/10.1142/S0219843614500224

Alemi, M., Meghdari, A., & Ghazisaedy, M. (2015). The Impact of social robotics on L2 learners' anxiety and attitude in English vocabulary acquisition. *International Journal of Social Robotics*, *7*(4), 523-535.

Batane, T., & Ngwako, A. (2017). Technology use by pre-service teachers during teaching practice: Are new teachers embracing technology right away in their first teaching experience? *Australasian Journal of Educational Technology*, *33*(1). https://doi.org/10.14742/ajet.2299

Becker, H. J. (1994). How exemplary computer-using teachers differ from other teachers: Implications for realizing the potential of computers in schools. *Journal of Research on Computing in Education*, *26*, 291–321.

Benitti, F. B. V. (2012). Exploring the educational potential of robotics in schools: A Systematic review. *Computers & Education*, *58*(3), 978–988.

Chamot, A. U. (2004). Issues in language learning strategy research and teaching. *Electronic Journal of Foreign Language Teaching*, *1*(1), 14-26.

Chao, K. J., Huang, H. W., Fang, W. C., & Chen, N. S. (2013). Embodied play to learn: Exploring K inect-facilitated memory performance. *British Journal of Educational Technology, 5*(44), E151-E155. https://doi.org/10.1111/bjet.12018

Chen, N.-S., Quadir, B., & Teng, D. C. (2011). Integrating book, digital content and robot for enhancing elementary school students' learning of English. *Australasian Journal of Educational Technology*, *27*(3). https://doi.org/10.14742/ajet.960

Cheng, Y. W., Sun, P. C., & Chen, N. S. (2018). The Essential applications of educational robot: Requirement analysis from the perspectives of experts, researchers and instructors. *Computers & Education*, *126*, 399-416.

Cheng, Y. W., Wang, Y., Yang, Y. F., Yang, Z. K., & Chen, N. S. (2021). Designing an authoring system of robots and IoT-based toys for EFL teaching and learning. *Computer Assisted Language Learning*, *34*(1-2), 6-34. https://doi.org/10.1080/09588221.2020.1799823

Cuban, L. (2001). *Oversold and underused: Computers in the classroom*. Harvard University Press.

DeKeyser, R. M. (2005). What makes learning second-language grammar difficult? A Review of issues. *Language learning*, *55*(S1), 1-25.

Domínguez, F., & Ochoa, X. (2017). Smart objects in education: An Early survey to assess opportunities and challenges. In *4th International Conference on eDemocracy & eGovernment-ICEDEG 2017* (pp. 216-220). IEEE. https://doi.org/10.1109/ICEDEG.2017.7962537

Ertmer, P. A., & Ottenbreit-Leftwich, A. (2013). Removing obstacles to the pedagogical changes required by Jonassen's vision of authentic technology-enabled learning. *Computers & Education*, *64*, 175-182.

Gul, S., Asif, M., Ahmad, S., Yasir, M., Majid, M., & Malik, M. (2017). A Survey on role of internet of things in education. *International Journal of Computer Science and Network Security*, *17*(5), 159-165.

Hadley, M., & Sheingold, K. (1993). Commonalties and distinctive patterns in teachers' integration of computers. *American Journal of Education*, *101*, 261-315.

Howland, J. L., Jonassen, D., & Marra, R. M. (2012). *Meaningful learning with technology* (4th ed.). Pearson.

Hubbard, P. (1992). A Methodological framework for CALL courseware development. In *Computers in Applied Linguistics: An International Perspective* (pp. 39-65). Multilingual Matters.

Hubbard, P. (2017). Technologies for teaching and learning L2 listening. In *The Handbook of Technology and Second Language Teaching and Learning* (pp. 93-106). https://doi.org/10.1002/9781118914069.ch7

Hung, I., Hsu, H., Chen, N., & Kinshuk. (2015). Communicating through body: A Situated embodiment-based strategy with flag semaphore for procedural knowledge construction. *Educational Technology Research and Development*, *63*, 749–769.

Hyun, E., Yoon, H., & Son, S. (2010). Relationships between user experiences and children's perceptions of the education robot. In *Proceeding of the 5th ACM/IEEE international conference on human-robot interaction* (pp. 199–200). ACM.

Janík, T., Seidel, T., & Najvar, P. (2009). Introduction: On the power of video studies in investigating teaching and learning. In T. Janik, & T. Seidel (Eds.), *The Power of Video Studies in Investigating Teaching and Learning in the Classroom* (pp. 7-22). Waxmann.

Jonassen, D. H. (1996). *Computers in the classroom: Mindtools for critical thinking*. Prentice-Hall, Inc.

Kanda, T., Hirano, T., Eaton, D., & Ishiguro, H. (2004). Interactive robots as social partners and peer tutors for children: A Field trial. *Human–Computer Interaction*, *19*(1), 61-84.

Kanda, T., Sato, R., Saiwaki, N., & Ishiguro, H. (2007). A Two-month field trial in an elementary school for long-term human-robot interaction. *IEEE Transactions on Robotics*, *23*(5), 962-971.

Karim, M. E., Lemaignan, S., & Mondada, F. (2015). A Review: Can robots reshape K-12 STEM education? In *2015 IEEE international workshop on advanced robotics and its social impacts (ARSO)*. IEEE. https://doi.org/10.1109/ARSO.2015.7428217

Khaddage, F., Müller, W., & Flintoff, K. (2016). Advancing mobile learning in formal and informal settings via mobile app technology: Where to from here, and how? *Educational Technology & Society*, *19*(3), 16-26.

Kim, Y., & Baylor, A. L. (2006). A Social-cognitive framework for pedagogical agents as learning companions. *Educational Technology Research and Development*, *54*, 569–596.

Kirsh, D. (2013). Embodied cognition and the magical future of interaction design. *ACM Transactions on Computer-Human Interaction*, *20*(1),1-30. https://doi.org/10.1145/2442106.2442109

Kozima, H., Michalowski, M., & Nakagawa, C. (2009). A Playful robot for research, therapy, and entertainment. *International Journal of Social Robotics*, *1*, 3-18

Leite, I., Martinho, C., & Paiva, A. (2013). Social robots for long-term interaction: A Survey. *International Journal of Social Robotics*, *5*(2), 291-308.

Leite, I., Martinho, C., Pereira, A., & Paiva, A. (2009). As time goes by: Long-term evaluation of social presence in robotic companions. In *RO-MAN 2009-the 18th IEEE international symposium on robot and human interactive communication* (pp. 669-674). IEEE. https://doi.org/10.1109/ROMAN.2009.5326256

Li, J. (2015). The Benefit of being physically present: A Survey of experimental works comparing copresent robots, telepresent robots and virtual agents. *International Journal of Human-Computer Studies*, *77*, 23–37.

Lin, V., Yeh, H. C., & Chen, N. S. (2022). A Systematic review on oral interactions in robot-assisted language learning. *Electronics*, *11*(2), 290. https://doi.org/10.3390/electronics11020290

Lincoln, Y., & Guba, E. (1985). *Naturalistic inquiry*. Sage.

Liu, D.-J., Huang, R.-H., Chen, N.-S., & Fan, L. (2016). *The Current situation and trend of global development of educational robots*. Posts & Telecommunications Press.

Madakam, S., Ramaswamy, R., & Tripathi, S. (2015). Internet of Things (IoT): A Literature review. *Journal of Computer and Communications*, 3, 164-173. http://doi.org/10.4236/jcc.2015.35021

Morales, S. (2017). Locating the European digital kitchen in its research context. In *Task-based Language Learning in a Real-World Digital Environment: The European Digital Kitchen* (pp. 19-42). Bloomsbury Publishing

Moreno, R., & Mayer, R. (2007). Interactive multimodal learning environments. *Educational Psychology Review*, *19*, 309-326.

Mubin, O., Stevens, C. J., Shahid, S., Al Mahmud, A., & Dong, J.-J. (2013). A Review of the applicability of robots in education. *Journal of Technology in Education and Learning*, *1*(209–0015), 13. http://doi.org/10.2316/Journal.209.2013.1.209-0015

Nikolaevskaya, I. (2017). Multidimensional model of the ZPD as tool of analysis of the child's cognitive-personal dynamics of development while overcoming learning difficulties. *CRIRES International Review: Innovating in the Vygotsky Tradition, 4*(1), 154-160.

Pachidis, T., Vrochidou, E., Kaburlasos, V., Kostova, S., Bonkovic, M., & Papić, V. (2019). Social robotics in education: State-of-the-art and directions. In *International Conference on Robotics in Alpe-Adria Danube Region* (pp. 689-700). Springer.

Pallotti, G., Niemants, N., & Seedhouse, P. (2017). Vocabulary learning in a real-world digital environment. In *Task-based Language Learning in a Real-World Digital Environment: The European Digital Kitchen* (pp. 207-229). Bloomsbury Publishing

Pasfield-Neofitou, S., Huang, H., & Grant, S. (2015). Lost in second life: Virtual embodiment and language learning via multimodal communication. *Educational Technology Research and Development*, *63*, 709–726. https://doi.org/10.1007/s11423-015-9384-7

Pica, T., Kang, H. S., & Sauro, S. (2006). Information gap tasks: Their multiple roles and contributions to interaction research methodology. *Studies in Second Language Acquisition*, *28*(2), 301-338.

Ramlowat, D. D., & Pattanayak, B. K. (2019). Exploring the Internet of Things (IoT) in education: A Review. In S. Satapathy, V. Bhateja, R. Somanah, X. S. Yang, & R. Senkerik (Eds.), *Information Systems Design and Intelligent Applications*. *Advances in Intelligent Systems and Computing* (Vol. 863). Springer. https://doi.org/10.1007/978-981-13-3338-5_23

Roos, J., Reshöft, N., Hartung, L., & Bußwinkel, J. (2017). Assessing and promoting language development in an interactive learning environment. *Task-based Language Learning in a Real-World Digital Environment: The European Digital Kitchen*, 137-156.

Seedhouse, P. (Ed.). (2017). *Task-based language learning in a real-world digital environment: The European digital kitchen* (Vol. 4). Bloomsbury Publishing.

Shilling, C. (2017). Body pedagogics: Embodiment, cognition and cultural transmission. *Sociology*, *51*(6), 1205-1221.

Skulmowski, A., Pradel, S., Kühnert, T., Brunnett, G., & Rey, G. D. (2016). Embodied learning using a tangible user interface: The Effects of haptic perception and selective pointing on a spatial learning task. *Computers & Education, 92-93*, 64-75.

Spolaôr, N., & Benitti, F. B. V. (2017). Robotics applications grounded in learning theories on tertiary education: A Systematic review. *Computers & Education*, *112*, 97–107.

Strauss, A., & Corbin, J. (1990). *Basics of qualitative research: Grounded theory and procedures and techniques*. Sage.

Tanaka, F., Cicourel, A., & Movellan, J. (2007). Socialization between toddlers and robots at an early childhood education center. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(46),17954-17958.

Toh, L. P. E., Causo, A., Tzuo, P.-W., Chen, I., & Yeo, S. H. (2016). A Review on the use of robots in education and young children. *Educational Technology & Society*, *19*(2), 148-163.

Tondeur, J., Kershaw, L. H., R. Vanderlinde, R., & van Braak, J. (2013). Getting inside the black box of technology integration in education: Teachers' stimulated recall of classroom observations. *Australasian Journal of Educational Technology*, *29*(3). https://doi.org/10.14742/ajet.16

Vygotsky, L. S. (1962). *Thought and language* (E. Hanf-mann & G. Vakar, Trans.) MIT Press.

Vygotsky, L. S., & Cole, M. (1978). *Mind in society: Development of higher psychological processes*. Harvard university press.

Wells, G. (2001). The Case for dialogic inquiry. In G. Wells (Ed.), *Action, talk, and text: Learning and teaching through inquiry* (pp. 171-194). Teachers College Press.

Wood, D., Bruner, J. S., & Ross, G. (1976). The Role of tutoring in problem solving. *Journal of Child Psychology and Psychiatry and Allied Disciplines, 17*, 89-100.

Xie, L., Antle, A. N., & Motamedi, N. (2008). Are tangible more fun? Comparing children's enjoyment and engagement using physical, graphical and tangible user interfaces. *Proceedings of the Second International Conference on Tangible and Embedded Interaction* (pp. 191-198).

Lin, R. P., & Wu, W.-C. V. (2023). Exploring Multiliteracy of Pre-Service Language Teachers through Spherical Video-Based Virtual Reality. *Educational Technology & Society*, *26*(3), 101-114. https://doi.org/10.30191/ETS.202307_26(3).0008

Exploring Multiliteracy of Pre-Service Language Teachers through Spherical Video-Based Virtual Reality

Rae Ping Lin[1] and Wen-Chi Vivian Wu[2,3*]

[1]Department of English Language, Literature, & Linguistics, Providence University, Taiwan // [2]Department of Foreign Languages and Literature, Asia University, Taiwan // [3]Department of Medical Research, China Medical University Hospital, China Medial University, Taiwan // raelin@gm.pu.edu.tw // vivwu123@asia.edu.tw

[*]Corresponding author

ABSTRACT: This qualitative study aims to explore how the production of language learning materials using spherical video-based virtual reality (SVVR) affords pre-service teachers multiliteracy development while also attempting to discover their perceptions toward adopting this emerging technology for future language teaching. Data from multiple sources was collected from pre-service English teachers enrolled in a TESOL graduate program in Taiwan, including video-recordings of the participants' presentations on their final SVVR projects and their self-generated VR teaching materials/artifacts, with two one-to-one semi-structured interviews further analyzed based on thematic analysis. The major findings demonstrate that through conducting the SVVR project, participants learned: (1) to compose multimodal lessons; (2) to concretize intangible contexts for learning; and (3) to use space as a mode for teaching and learning. The authors presented insights into affordance of SVVR material production for their multiliteracy development of engaging pre-service English teachers, as well as their perceptions with regard to this SVVR hands-on experience. Moreover, the authors offer recommendations for putting such experience into practice.

Keywords: SVVR, Multimodality, Pre-service English teachers

1. Introduction

With its characteristics of being interactive and immersive within a realistic learning environment, virtual reality (VR) has been widely applied in the language learning context and has shown great potential for enhancing L2 learning achievement (e.g., Ebadi & Ebadijalal, 2020). VR refers to computer technology that can situate users in real-world simulated environments to stimulate interaction between users and virtual objects in a virtual context (Ghanbarzadeh et al., 2014). The two main types of VR are identified: low-immersive VR (LiVR) and high-immersive VR (HiVR). While LiVR displays three-dimensional (3D) virtual space experiences through conventional computer devices, (e.g., desktop computers), HiVR presents a 360° 3D virtual context that creates a sensation of physical presence provided by head-mounted displays (HMDs) (Kaplan-Rakowski & Gruber, 2019). LiVR is typically more accessible and cost-effective, but appears less realistic to the senses. On the other hand, HiVR creates a great sense of immersion and high interaction but is often more costly and necessitates more skill to use than LiVR. (Chang et al., 2020). Recently, spherical video-based virtual reality (SVVR) has emerged, a new VR form that can rely on either LiVR or HiVR and is designed to be low-cost, easy-to-access and easy-to-produce (Jong et al., 2020). SVVR delivers 360-degree spherical videos for viewers to look in all directions by rotating what they want to see with an HMD or with a conventional computer device (Chen & Hwang, 2020). Given its features of financial affordability and its relative ease-of-use with regards to computer skills, language instructors are able to design and improve their own teaching materials and activities for creating an immersive and interactive environment for students to learn a second language (Chien et al., 2020).

In this era of switching from printed materials to the screen, McVee et al. (2008) posited that technology should not be viewed simply as hardware and software, in effect a tool for enhancing learning. Instead, technology is about literacy, a process that synergizes semiotic resources for people to create meanings. This transactional stance (Bruce, 1997) emphasizing the dynamic and interwoven relationship of technology and literacy resonates with the New London Group's new literacy. According to the New London Group (1996), communication is not limited to written or spoken interaction, but expands itself to multimodal forms of expressing oneself by combining linguistic, visual, audio, gestural, and spatial texts afforded by digital and internet technology. Therefore, modern literacy should shift away from traditional literacy with its primary focus on writing and speaking toward multimodal literacy, defined as "the meaning-making skills constituting reading, viewing, understanding, responding to and producing and interacting with multimedia and digital texts" (Walsh, 2010, p. 213). According to Rowsell and Walsh (2011), multimodal literacy focuses on "multimodality," the ability to use different modes to create meaning. In contrast, "multiliteracy" refers to the pedagogical potential of incorporating the transformational new form of literacy into classrooms to engage students so as avail themselves to the multiple forms of digital technology.

Regarding language education, Veliz and Hossein (2020) have also pointed out that in the multifaceted technology-mediated world, it is almost inevitable to incorporate digital multimodal tools into language classrooms for enhancing language learning experiences. Language teachers' multiliteracies have thus become essential for teaching a second language in the modern digital age (Jiang et al., 2021). Li (2020) argues that while an increasing number of studies have documented exactly how language learners engage in, and benefit from, a wide range of multimodal practices, research on multimodal experiences of language teachers is still in its infancy. This leaves teachers feeling unprepared, and they therefore hesitate to incorporate multimodal practices into their classrooms. Considering that language teachers' multimodal experiences and knowledge play an important role in students' multimodal literacy development, aligning with Rowsell and Walsh's (2011) concept of "multiliteracy," this study is designed to explore how pre-service teachers' multiliteracies develop through utilizing SVVR tools to design teaching materials in an English as a Foreign Language (EFL) context. The following questions guide us through this inquiry:

- RQ1: How does a SVVR project afford pre-service teachers' multiliteracy development in English language teaching?
- RQ2: What are the perspectives of pre-service teachers toward using SVVR to enhance their English language teaching?

Given that teacher preparation is fundamental in implementing new technology with multimodal-informed pedagogy, this study is significant. It offers opportunities for SVVR technology to be introduced to, and practiced by, pre-service teachers while also exploring their multimodal pedagogy development and their perceptions toward adopting SVVR for language teaching.

2. Literature review

2.1. From digital literacy to multimodal literacy in language teacher education

The ability to understand and utilize technology has been coined "digital literacy" (Dudeney & Hockly, 2013), a term often used interchangeably for new media literacies, twenty-first century skills, or digital competencies (Jeong, 2017). As teachers play a pivotal role in transforming technology into meaningful language learning, numerous studies have investigated the integration of digital literacy into language teacher education (Kuru-Gönen, 2019; Jeong, 2017). However, as Dzekoe (2020) rightly pointed out, in an age when people use technology not merely as a tool, but as a process to design semiotic products and events, almost all screen-based texts are multimodal in nature. Our understanding of using technology in language teaching settings should go beyond digital literacy, with its focus mainly on the ability to utilizing different technology tools. As such, it can move toward a "multimodal literacy" that captures learners' understanding of the multimodal affordance of different digital tools: that is, how digital technologies are designed with different modes and what they allow users to produce and achieve so as to express meaning (Dzekoe, 2020). Accordingly, language teachers' multiliteracies—the ability to integrate multimodality into classrooms—have become essential in training their students to become designers able to use a wide-ranging multimodal repertoire to "interpret, manage, share, and create meaning in the growing range of digital communication channels" (Dudeney et al., 2013, p. 2).

In L1 teacher education programs for English Language Art (ELA), Miller (2007) and Hundley and Holbrook (2013) had pre- and in-service teachers conduct a multimodal composition project wherein they combined different modes to design multimodal materials for teaching the English language. Teachers in both studies revealed that they gradually came to see themselves as an active creator who orchestrated visuals, music and printed text in the project to transform their students' learning experiences from reading/writing print to visual/auditory texts. However, some teachers still hold a "print bias" (Miller, 2007, p. 64) and consider multimodal texts to be illegitimate instructional materials, and therefore showed reluctance to teach English language in multimodal formats. In an L2 learning context, both Rance-Roney (2010) and Li (2020) reported that the participating teachers in their graduate courses showed very positive benefits in developing their professionalism in teaching English language through engaging in multimodal practices using technological platforms of their own preference (e.g., Prezi and Storybird). However, Yi and Choi's study (2015) echoed the teachers' experiences in Miller's work (2007) in which, though acknowledging the potential of multimodal pedagogy, the in-service teachers in their study showed concern that too much time spent on various modes other than text might hurt the academic performance of their students, as learning outcomes were still assessed primarily through text mode. Yi and Choi (2015) called for more investigation into teacher education about multimodal pedagogy to develop the ability of pre-/in-service teachers to design proper multimodal lessons which align with mandatory standards in their teaching contexts.

2.2. VR and SVVR for language learning and teacher education

With the characteristics of being interactive and immersive within a realistic learning environment, various VR channels (e.g., Second Life, Google Expeditions) have offered great potential for enhancing L2 learning achievement (Chen et al., 2021; Ebadi & Ebadijalal, 2020; Wu et al., 2021), while providing ample opportunities for students to experience contexts and concepts which might not be viable in the real world (Choi et al., 2016). Scholarly attention has also been increasingly geared toward implementing VR into language teacher education. For instance, Chen (2019) showcased how an ESL teacher, through the guidance of a mentor, learned how to develop English lessons using Second Life to motivate students in a remedial English support program. Moreover, Kozlova and Priven (2015) investigated how a group of pre-service teachers in Canada used 3D virtual worlds to design engaging English language lessons for undergraduate EFL students in Turkey.

Despite the pedagogical benefits offered by VR channels, they are often presented in fixed sets of images which cannot be adjusted by teachers for developing more flexible teaching materials. Moreover, it is usually technically and financially challenging for teachers and schools to develop realistic 3D interactive VR content (McFaul & FitzGerald, 2019). Even when a developed VR platform is available, it is difficult for most teachers to learn to design VR content for pedagogical purposes when no sufficient instruction is provided (Chen, 2019). Recently, spherical video-based virtual reality (SVVR) has been used to address these difficulties. SVVR refers to a VR approach presenting 360-degree spherical videos, creating a realistic environment for users to look in all directions (Chen & Hwang, 2020; Walshe & Driver, 2019) with the use of an HMD or conventional computer device. Compared to the general HiVR, SVVR can be easily and inexpensively implemented in classrooms, as it offers a variety of ready-to-use 3D spherical videos and objects for users to design their own VR tasks. Users can also create their own spherical 360-degree videos with a handy spherical (360-degree) camera or download free available spherical videos online. Secondly, SVVR can be accessed by users' own mobile phones using an inexpensive cardboard goggle setup, costing as low as US$6 (Jong et al., 2020); alternatively, it can be viewed without goggles from standard computer devices.

SVVR has been recognized for its features of being low-cost, easy-to-access and easy-to-produce, and has thus been widely applied for educational purposes (Han, 2020). Only a recent handful studies have focused on using SVVR in language education. For example, Chen et al. (2021) examined the effect of integrating SVVR technology into problem-based English learning of Taiwanese engineering majors, finding a positive effect on their English vocabulary acquisition and motivation toward future English learning. Moreover, Chen and Hwang (2020) found significant improvement in English oral presentation ability and learning motivation in students having different cognitive styles (i.e., field-dependent and field-independent) who experienced SVVR learning in an English for Tourism course. Lastly, Chien et al. (2020) implemented peer assessment into SVVR-based English courses in Taiwanese high schools. The results showed that the peer-assessment-based SVVR group revealed significant improvement in English speaking, learning motivation, critical thinking skills, and reduction of anxiety toward learning English.

As shown in this section, many researchers have noticed the potential of VR and SVVR in educational applications and language education. Yet research on teacher training in using VR, particularly SVVR, is still relatively scarce. Moreover, little attention has been paid to multimodality or multiliteracy development in a VR context. According to Philippe et al. (2020), there were no peer-reviewed articles found published between 2010 and 2020 on Science Direct (Elsevier) related to multimodality, VR and teaching and learning. Given that the teacher plays a central role in developing students' multimodal literacy when utilizing a new technology, for example VR, teachers' multiliteracies in the context of VR application deserves more examination.

3. Methodology

3.1. Context and participants

Thirteen graduate students majoring in English enrolled in an 18-week elective course entitled Teaching Materials and Pedagogy in TESOL at a four-year academic university in central Taiwan. Seven of the thirteen students volunteered to participate in this study: among the seven participants, while five of them agreed to be interviewed and to share their SVVR artifacts for this research, another two agreed to be interviewed but had no intention to share their artifacts for research purposes due to personal privacy issues. Consent forms were signed by all participants before data collection. Ages ranged between 22 and 31. Most of these pre-service English teachers had little or no formal teaching experience prior to the class, and none of them had experienced VR instruction in their previous courses or used the SVVR technique to create learning materials of their own.

However, all of them had used VR for their own entertainment purposes, such as movies, tourism, museums, or gaming. The students learned about general TESOL teaching principles for the first nine weeks, after which each student needed to design a lesson plan in the English language on a topic of their interest (Wks. 10-12). Students were then introduced to an SVVR tool, Uptale (Wk. 13), and asked to design an individual SVVR project based on the lesson plan as their final project (Wks. 14-17). At the end of the semester, students had to do an individual presentation and demonstrate their SVVR projects to the class (Wk. 18).

3.2. Production software and technology used for SVVR learning materials production

The SVVR learning materials produced by the pre-service teachers were completed through Uptale (https://www.uptale.io), a web-based authorized platform accessed by computer for the building of an immersive, contextualized, and interactive learning environment. The researchers chose Uptale due to its affordable cost and its relative ease-of-use regarding necessary computer literacy/skills. Uptale allows users to attach 360-degree panoramic videos/images along with text descriptions, narratives, questions for interaction, ambient sounds and voice recording to each scene. The 360-degree panoramic videos are available from Uptale or can be captured using 360-degree cameras by users. To ensure that the pre-service teachers would understand and make full use of the available features of Uptale, the course instructor (the second author) provided a three-hour training workshop in advance wherein the main features of Uptale were introduced. Familiarized with Uptale features, the participants then arranged the scenes contextually based on their pre-designed lesson plans, which included the main goals of the class, the level of its learners, learning focus, learning time and activities. To conclude, through Uptale SVVR, the hypothetical students of the participants in this study could learn in a more immersive and interactive environment through 360-degree spherical images or videos that are more realistic than 3D animations.

3.3. Research design

The researchers adopted a qualitative research design to address the research questions (Creswell & Poth, 2016). The researchers collected data from three sources: (1) participants' artifacts from the SVVR project (n = 5); (2) video-recordings of the participants' presentations about their final SVVR artifacts (n = 5), each presentation recording lasting approximately 25 minutes, including a Q&A section; and (3) two semi-structured one-on-one interviews with each volunteer interviewee (n = 7), the first interview conducted one day after their final presentations and the other after initial analysis for follow-up questions and member-checking. Each interview lasted one hour and was conducted by the first author after the final grades for the course were released. Due to circumstances surrounding the pandemic at the end of research course, participants' presentations and research interviews were conducted and recorded online through Microsoft Teams software. While the final SVVR presentations were conducted in English, individual interviews were in Mandarin. All data sets were used to answer RQ1, while interview data was used to answer RQ2.

3.4. Data analysis

For the presentation and interview data, we adopted Braun and Clarke's (2006) thematic analysis method and followed the six steps they suggested which allowed the researcher to identify, analyze, and report patterns across data sets to address the research questions appropriately and adequately (Table 1).

As for SVVR artifacts, we followed Bezemer and Kress (2016) as our analytical guide and understood that communication, including teaching and learning, is a process of meaning-making. We analyzed the creation of the SVVR artifacts as a process of how the pre-service teachers, the meaning-makers, used a variety of modes available on Uptale tool as meaning-making resources to "construct social (learning) environment" meaningful for their students (p. 62). Along with the rationales provided by the participants in the presentation and interview data for using different modes to design their SVVR teaching materials, we better understood how the SVVR project afforded their multiliteracy development. During the analysis course, we constantly compared and contrasted the relevant features among the three major data sources—SVVR artifacts, presentation recordings, and interviews—to triangulate and enrich the dataset in order to reach a sound understanding and interpretation of the participants' experiences and multiliteracy development in producing SVVR projects for English language teaching. During the process of producing the report, the Mandarin transcription was translated by the first researcher into English, then checked by second researcher and the pre-service teachers before being presented as quotes in this paper. Pseudonyms were used to conceal the identities of the participants. For RQ2, all responses of the volunteer interviewees (n = 7) will be presented in the paper. For RQ1, the researchers presented an analysis of the five SVVR projects from the interviewees who were willing to share their artifacts and respective comments.

Table 1. Phases of thematic analysis

Phase	Description of the process
0. Transcribing data	Transcribing the spoken data in presentations and interviews verbatim in English or Mandarin.
1. Familiarizing yourself with your data	The first author read and re-read the data along with corresponding SVVR artifacts, noting initial ideas.
2. Generating initial codes	The first author coded relevant features in relation to research questions across multiple data sets.
3. Searching for themes	The first author collated codes repeatedly and identified the themes in relation to the participants' experiences of conducting the SVVR project and multiliteracy development.
4. Reviewing themes	Based on the emerging themes categorized by the first author, two researchers reviewed and discussed the themes together.
5. Defining and naming themes	Two researchers re-coded the data and re-organized the themes until they could clearly define and name the themes.
6. Producing the report	Two researchers selected vivid and compelling excerpts for each theme to answer the research questions and finally produced a scholarly report of the analysis.

4. Findings and discussion

4.1. Pre-service teachers' multiliteracy development in English language teaching

4.1.1. Learning to compose multimodal lessons

Several participants reported that they were worried that they might not be able to complete the SVVR tasks because of their limited technology skills. However, they also mentioned that through doing this SVVR project, they became more confident and motivated to design English lessons that differ from the traditional way of teaching. As Cindy commented, through this SVVR project she realized that:

> Traditional learning methods such as memorizing vocabulary and grammar rules from a textbook cannot really motivate students to learn the English language in today's digital world. (Interview)

Therefore, she was motivated to design a task for students to learn vocabulary of different animals in the grasslands of Africa (see Figure 1).

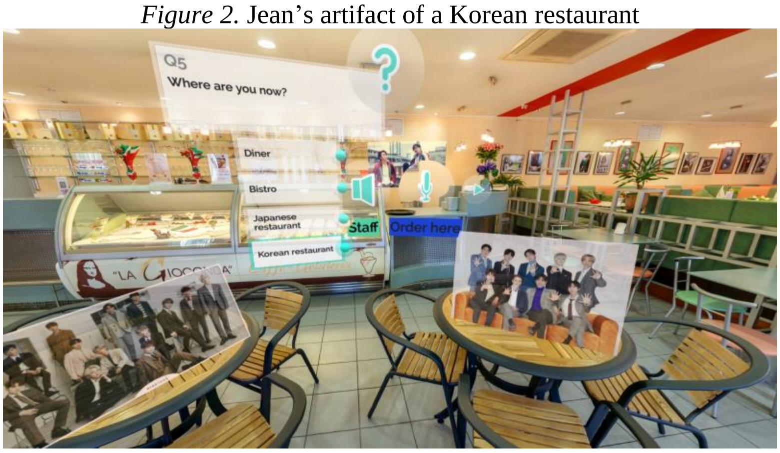
Figure 1. Cindy's artifact of African grasslands

Cindy designed a lesson on the SVVR platform integrating modes such as text (questions and answer options), sounds (reading out-loud of questions and answer options) and African folk music along with a 360-degree spherical video of the African grasslands created by herself. By composing a task with these modes, Cindy wanted her students “to feel learning is actually interesting and fun, just like playing games” (Presentation). Cindy later commented that she particularly liked the spherical view because “students were really immersed in and surrounded by the African grasslands while they played” (Post-interview). Compared to learning from printed teaching materials, Cindy believed that it could be easier to learn the vocabulary of African animals when students were surrounded by a virtual environment where they could explore the 360-degree image with folk music playing. As she excitingly said, “it’s just like they were sitting in a jeep, driving around and exploring!” (Post-interview).

Jean emphasized the importance of the spoken mode made possible in this tool. For example, Jean designed a task where a mom called her daughter (the students) to order a take-out meal (Figure 2). First, students had to listen carefully to what they needed to get for the mom. They then pressed the button and a Korean staff member took their orders. The students then pressed the button and ordered the food as the mom required. Their voices were recorded, and students could press the button to hear their own voice played back. Jean stated:

I think it is good to have students be able to interact and speak up to the staff. It is easy for them to get all kinds of input, but rarely they can practice speaking in a space like this…they can learn all the aspects here. (Interview)

Figure 2. Jean’s artifact of a Korean restaurant



Jean observed that despite being constantly exposed to a digital environment, digital natives seldom had the opportunity to speak up in that space. The interactive environment afforded by the SVVR tool offered hypothetical students a valuable learning space to learn all aspects of the English language, including the skills of interacting and speaking with others. In addition to the spoken mode, Jean also integrated pictures of a Korean idol group throughout the 360-degree spherical video with Korean pop songs playing. Jean believed that the images and music could contextualize the interactional environment and make the space more authentic for interaction.

While several participants had noticed the importance of integrating a variety of modes into their teaching material, Jill noted:

> I was afraid that I couldn’t express myself clearly through different ways, so I still relied on text to go with the sounds and pictures to explain what I wanted them to do. (Presentation)

Jill’s experience demonstrated that along with various semiotic means, such as visual and aural, text mode also serves as equally important to other semiotic modes for people to express themselves (McVee et al., 2008). As a novice teacher, Jill found the traditional text modes offered her a sense of security to clearly express her teaching purposes to her future students. According to Early et al. (2015), written and spoken modes are also important means for communication in our new age and therefore should not be ignored. Enabled by its voice recording and voice recognition functions, the SVVR tool used in this study afforded an interactive space for pre-service teachers to include visual and audio (Cindy), speaking (Jean) and written forms (Jill) as modes to enrich their multimodal teaching tasks.

4.1.2. Concretizing the intangible context for learning

Several participants also showed the ability to compose their VR instructional artifacts, using different modes, into a highly immersive learning environment for students to experience contexts that might not be accessible in their real world (Choi et al., 2016). For instance, aware that her students might not be able to physically view the great migration in Africa, Cindy created a 360-degree spherical image coupled with a video showing the Great Migration of African antelopes to introduce the concept and vocabulary "Great Migration" to students (Figure 3). As she explained:

> Using only sounds or pictures can't really show how fast antelopes are moving in such a big crowd when they migrate. I think the video can show the speed and sounds of antelopes, even the flow of water. It can immerse them into the real context, like the Great Migration is really happening around them. (Interview)

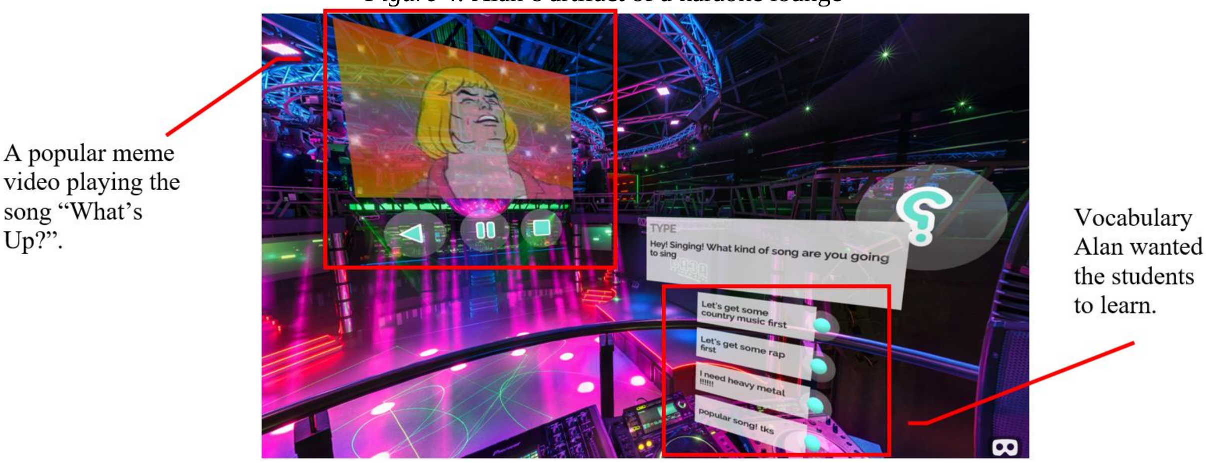
Figure 3. Cindy's artifact of the Great Migration

With a 360-degree image and a video of the Great Migration, Cindy used multiple modes, such as images of the animals, motion (e.g., speed), and the sounds of running antelopes and water flow from the video to simulate the great migration. By doing so, she hoped that students could learn the concept and vocabulary while "it is really happening around them."

Similarly, for high school students to experience singing in a karaoke lounge, a place where the majority of students might not have the chance to go because of the cost, Alan created a SVVR project for students to experience this activity while learning English:

> I couldn't find an image of a karaoke lounge from the tool so I chose this place with dark but colorful lighting…I put on a very popular music video to make it more like a karaoke place and also to catch the students' attention for learning. (Presentation)

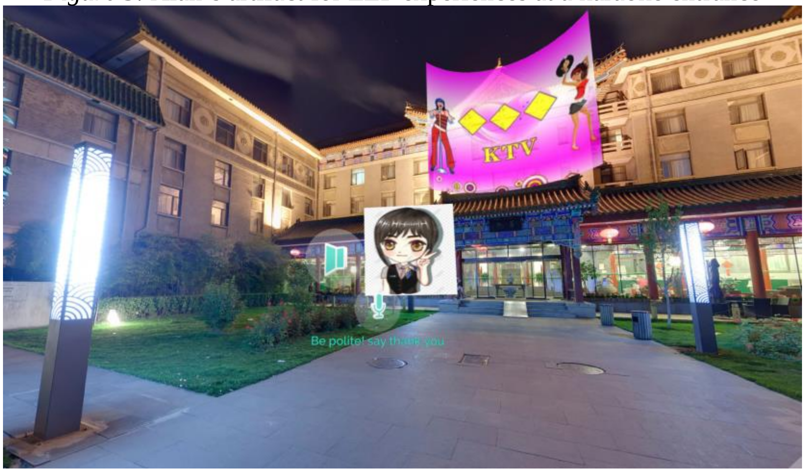
Figure 4. Alan's artifact of a karaoke lounge



In this scene (Figure 4), Alan chose a background picture with "dark but colorful lighting," where he put on a popular music video among young people with the song playing "What's Up" by 4 Non-Blondes. The music video was not used for karaoke singing, but was commonly used among young people as a meme asking for information on social media. With its colorful lighting and popular music video, Alan orchestrated a multimodal environment that was close to the students' life and culture, in order to trigger their interest in learning English vocabulary such as country music, rap, and heavy metal.

What is also worth noting was that several participants even created an intercultural communication context for their students to experience English as a lingua franca (ELF) by adding different English accents in the SVVR task. For example, Alan purposely added a Japanese accent in one conversation in one karaoke scene as shown in Figure 5.

Figure 5. Alan's artifact for ELF experiences at a karaoke entrance



At the entrance, a waitress welcomed the customers by saying, "There are still rooms available. Please check in with the counter." By pushing the button, the students answered "thank you" in order to move on to the next scene. Among many other accents available from the SVVR tool, including native English speakers, Alan chose a Japanese accent for the waitress because:

> It is very likely that people will encounter English speakers from all around the world…it would be good to train the students to understand different kinds of accents." (Presentation)

In this globalized world, English is used as a means of interaction. Meaning should come foremost and forms second. From this perspective, different accents and varieties of English should be valued for intercultural communication (Jenkins, 2009). By orchestrating an artifact with a 3D picture of a local karaoke place and an image of a Japanese waitress speaking with a Japanese accent, Alan created an ELF space for his students in Taiwan so as to learn to accommodate Japanese accents in English-mediated interaction.

Like Alan, Jean also created a similar space (see Figure 2) by choosing a Korean-English accent to represent the staff's voice in order to make students feel that they were "really situated in a Korean restaurant" (Post-interview). As English has become a lingua franca, the language has been globalized into many forms and accents for communication in various interactional context (Jenkins, 2009). The SVVR tool allows multiple modes to orchestrate such a globalized space in the virtual world for Alan, Jean and their students to experience a globalized context of English communication.

4.1.3. Learning "space" as a mode for teaching and learning

"Space" is an important piece of the multimodal repertoire for communication (Keating, 2016). In the highly immersive and interactive SVVR project, several participants realized that space was not merely a vessel, but text, that is, meanings people design and compose to communicate for their own purposes (Bezemer & Kress, 2016). For example, Nicky chose an image with a big counter (Figure 6) as her scene background for students to learn how to tell clinic staff their appointment time with the doctor. When asked about her choice of such a big counter, she stated that the counter made the space more interactive for learning:

> The counter helps people know more precisely what situation they are in and what they need to do in that situation, such as needing to greet first and then explain the purpose for coming here. (Interview)

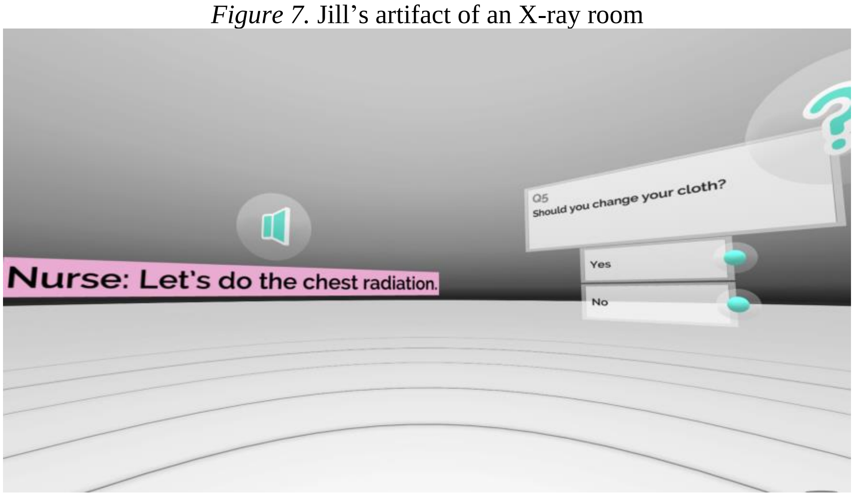
Figure 6. Nicky's artifact of a clinic counter

According to Jucker et al. (2018), interaction takes place in a spatial context when interactants become aware of being perceived. The counter space here does not refer to merely a physical object, but also to a semiotic resource Nicky used to help students become aware that they are situated at a clinic counter and within the space of a particular institutional discourse, like "you need to greet first and then explain your purpose for coming here." With the text instruction "tell the counter staff your appointment time with the dentist," students are made to notice that they are being perceived as an interactant in the space. Using the big counter along with the text instruction, Nicky constructed the space as a mode for students to visualize the conversation in which they had to interact with the other interactant (i.e., clinic staff), using language appropriate for a dentist clinic counter.

Likewise, Jill chose an image of an X-ray room (Figure 7) as a scene background and constructed a pedagogical space for students to learn how to interact by "feeling" the X-ray room. The following quote relates her description of the scene design:

> This is how I felt when I was in an X-ray room. The room is grey and looks spacious. The door is pretty heavy and the whole room is cold and mechanical. I felt people don't talk much but just want me to follow instructions…to finish and leave there as soon as possible. (Interview)

Figure 7. Jill's artifact of an X-ray room

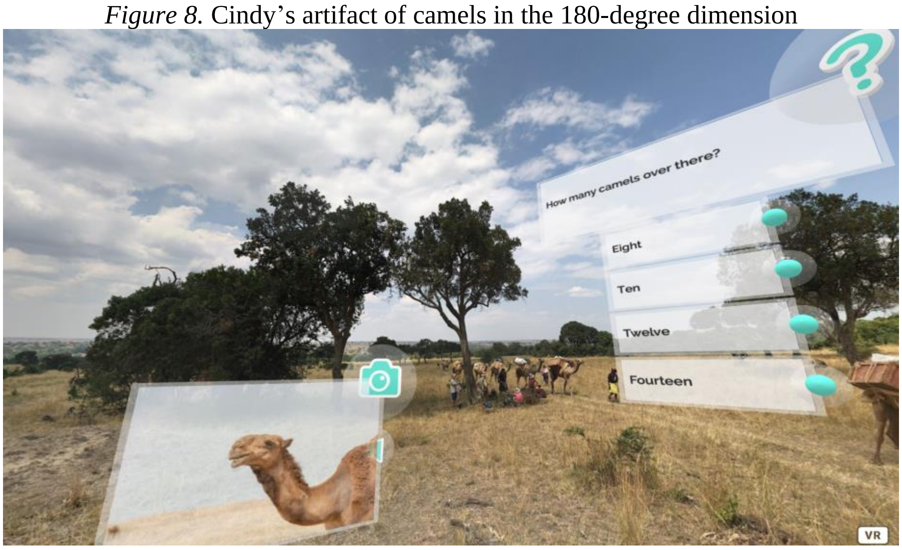


Using the color "grey," Jill depicted the X-ray room as "cold, heavy, and mechanical" that make people "feel alone." According to Jucker et al. (2018), senses serve as a central role in constructing our experiences and understanding of the world. By using her senses (e.g., cold, heavy, mechanical and alone) to construct the space, Jill constructed the space as a somewhat unpleasant place to stay, where minimum interpersonal interaction was required. Using the "grey color" to compose feelings in the space, Jill seemed to guide the students to understand the appropriate communication discourse in an X-ray room—that is, "no talking, just follow instructions." Jill further elaborated that situated in an unpleasant atmosphere, "the students will not linger around but finish the questions in this scene quickly and move on to the next scene for another task." (Interview). The space was presented as a text carrying communicative meanings between Jill and her students, instructing the students to 'finish the task quicky and move on to the next task."

Cindy also learned to organize the digital space to communicate with her students. She stated that "although it's a 360-degree spherical image, I tried to put the relevant clues for students within 180-degrees." As shown in Figure 8, there are only two clues in the 180-degree dimension; one is a picture of a camel, and the other is a question asking students to count the number of the camels appearing in the safari. In organizing the space in this manner, she hoped that her students could find the clues more efficiently:

> It is to have students learn English by answering questions through the clues that are given. If students are spending too much time finding the clues around the image, it will just be wasting time and not so effective. (Interview)

Figure 8. Cindy's artifact of camels in the 180-degree dimension



By placing relevant clues within the 180-degree dimension, Cindy tried to communicate with her students not to waste time searching all around the image, but rather focus on the dimension she had purposefully designed to include all the information for learning. The spherical image served as a mode for Cindy to communicate with her students about what needed to be learned in the space.

As Keating (2016) rightly pointed out, the digital world has given rise to new discourses about "how to appropriately organize digital spaces and how to interpret behaviors within them" (p. 264). While designing their SVVR artifacts, participants realized that the 3D VR space was not only a vessel that contained different learning tasks for students. Instead, using different modes available in the platform (e.g., big counter; grey color; dimensions), they organized their SVVR space appropriately as a pedagogical environment to communicate with their students about the learning foci in each of the digital scenes and to guide their students to interact appropriately in the context they designed.

4.2. Pre-service teachers' perspectives about integrating SVVR into English learning

4.2.1. Technology-novice friendly

Echoing many of the studies on using SVVR in an educational setting (e.g., Chien & Hwang, 2020), the majority of participants agreed that the SVVR tool required simple technology skills for designing VR tasks and was therefore technology-novice friendly. As Jean noted:

> I started learning how to incorporate new technologies to help my teaching… (this was) especially helpful for a technology-novice like me who only knows how to make simple PowerPoints" (Interview).

Through doing this project, Ava also stated:

> It took a bit of time to get used to each function, but after the workshop and discussion with Vivian, it was easier than I thought. I was pretty worried that I couldn't complete the task because my technology skills were really bad, but I did it! (Interview)

Ava's comment pointed out that not only can this SVVR tool be user-friendly, but also that sufficient instruction (e.g., workshop) can facilitate the pre-service teachers' learning experiences of integrating new technology into their teaching practices (Chen, 2019). The features of being easy-to-access and easy-to produce have not only made SVVR a friendly context for pre-service teachers to learn how to integrate VR tools into their own teaching, such features also trigger the motivation of teachers to use VR-related technologies in the future. As Nicky said:

> I used to think that these were not my things…but after doing this project, I think I became more comfortable and confident in trying and using 3D-related technology for my own purposes or for teaching purposes. (Interview)

Cindy also mentioned that with the step-by-step instruction provided by the course instructor, designing her own VR tasks was easier than expected. This smooth experience motivated her to design another teaching task using the Uptale SVVR tool, even after the course has finished. As she stated, "I am already planning to conduct another teaching task using the same platform." (Post-interview)

4.2.2. From passive to active learners

Given the feature of being interactive, several participants mentioned that the use of SVVR can turn not only their students but also themselves from passive learners to active learners. For example, Jill noted that in contrast to another technology medium—video which is often used as a pedagogical tool—SVVR transformed students from being passive to active learners as "they have to find the questions and interact with the computer, and it will respond back to them as well" (Interview). Similarly, Andy also remarked about the difference between watching a two-dimensional video and SVVR:

> SVVR is different from watching videos because students have to interact immediately with the speaker in the computer in order to get rewards and proceed to the next task. (Interview)

Andy further elaborated that by interacting to get rewards, the game-like activity was particularly efficient for motivating students to actively engage in English-language learning. In addition to encouraging students to actively react in the virtual space, Jean also noted that the virtual interactive environment in SVVR could make students less nervous when speaking English because they were not interacting with "a real person." As Jean commented:

> This makes it less stressful… You are talking to a virtual person and she is not really standing in front of you and waiting for you to speak up. I think this could motivate especially shy students to speak up. (Interview)

Although being interactive and immersive in the virtual space, the person with whom the students had to interact was not "really standing in front of you and waiting for you to speak up." This feature made the learning environment less intimidating and therefore had the potential to motivate shy students to speak out more actively.

Not only did the SVVR project provide a potential context to turn language students into active learners, both Cindy and Alan noted that in order to make their projects more immersive and authentic, they spent extra time after the class trying to learn how to make 3D-spherical videos by themselves or how to edit readily available videos they found online. As Cindy commented:

> Uptale does not have the safari picture I needed for my SVVR background, so I googled the topic and also asked my friends their advice as to how to make 3D-spherical videos for this project. The one I used was downloaded online. I learned to transfer it into a compatible form to put it in my SVVR project. (Interview)

Through completing this SVVR project, Cindy became an active learner who was willing to develop her own VR skills for her own interests or purposes of teaching the language.

4.2.3. More support needed from instructors and peers

Despite the many benefits expressed by the participants, a few drawbacks were also mentioned. First of all, Alan noticed that although 3D pictures were available, there were very limited options from within the SVVR platform itself. This limitation sometimes made the scene “inauthentic” or “not real” for the designed scenario. As Alan explained:

> I was trying to find a 3D picture of a waitress in a karaoke lounge, but there was no 3D portrait to choose from. So, I ended up putting a 2D picture there …as you can see, she does not really fit into the scene and looks very inauthentic. (Post-interview).

Nicky noticed the same issue. In order to render the scene a more authentic environment for the learning, she tried to make her own 3D pictures or 360-degree spherical video but found it very time-consuming and even costly. She doubted whether she would spend so much time on designing a teaching task for her future students. To address this issue, Nicky suggested that more workshops can be offered to demonstrate how teachers can make their own 3D images and spherical videos. With these skills, they could make the virtual learning space more immersive and effective for learning.

Andy also articulated that the time required was a significant challenge in designing the SVVR task. To address this issue, Andy suggested that it would be helpful to have other viewers to check the scenes for the designer. Having the perspectives of other viewers, Andy believed:

> They can tell me what needs to be removed and what needs to be added to improve the tasks so the students can learn more efficiently. (Interview)

Several participants also stressed the importance of having other viewers to provide feedback to improve their SVVR tasks. Ava said:

> Because I am the creator, the processes and structure of each scene all makes sense to me…This could be tricky because I wouldn’t know if I had made myself clear enough for my students to follow the instructions. (Interview)

To prevent such a blind spot, she had one classmate go through her task. The peer then suggested that some of Ava’s scenes were packed with too many questions and text, making the instructions not so clear to follow. Ava then revised some parts of her tasks accordingly. She concluded that peer feedback was very helpful as “it saved a lot of time for me and also made me feel more confident to present the task to others.” (Interview).

To conclude, although challenges were reported, participants also emphasized the importance of peer collaboration and the guidance of their instructor to help smooth the process of learning how to integrate technology such as SVVR into their multimodal teaching artifacts (Chen, 2019; Li, 2020).

5. Conclusion and recommendations

The findings show that pre-service teachers’ multiliteracies developed through conducting SVVR teaching projects. Learning by doing a SVVR project along with a training workshop and instructor’s guidance, the participants learned how to orchestrate their own SVVR teaching artifacts and created multimodality-informed materials to engage their students to learn the English language in an immersive, immediate, interactive, and multimodal learning environment (Chen, 2019). The ability of teachers to create such environments for language learning is crucial so as to expose their students to a wide range of multimodal repertoire, preparing them to be able to interpret, manage, and create meaning in the digital world (Dudeney et al., 2013). Teacher education thus plays an important role in transforming more teachers into creators of multimodal artifacts using new technologies such as SVVR (Li, 2020; Yi & Choi, 2015).

The multiliteracy concept, findings and discussion above lead to the following three recommendations for practice when integrating an SVVR tool to enhance the multiliteracy of pre-service teachers: (1) explicit and detailed instructions about the functions of SVVR tools is necessary to enhance pre-service teachers’ confidence about using new technology, as only when they feel comfortable and confident does the willingness to integrate technology to their own teaching increase; (2) technology skill levels differ regarding knowing how to use such tools to design language lessons, as guidance from instructors or peers is needed for pre-service teachers to learn how to integrate the new technology into the pedagogy of the language classroom; and (3) workshops for developing 3D images, 360-degree spherical videos or other modes (e.g., music editing) can be provided to train pre-service teachers to make their own virtual objects for SVVR teaching tasks, as when they are not limited to the object options available from the SVVR tool, they can create more powerful and effective SVVR multimodal teaching materials for their future students.

References

Bezemer, J. J., & G. R. Kress. (2016). *Multimodality, learning and communication: A Social semiotic frame.* Routledge, Taylor & Francis Group.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*(2), 77-101.

Bruce, B. C. (1997). Critical issues literacy technologies: What stance should we take? *Journal of Literacy Research*, *29*(2), 289–309. https://doi.org/10.1080/10862969709547959

Chang, S. C., Hsu, T. C., & Jong, M. S. Y. (2020). Integration of the peer assessment approach with a virtual reality design system for learning earth science. *Computers & Education*, *146*, 103758. https://doi.org/10.1016/j.compedu.2019.103758

Chen, C. H., Hung, H. T., & Yeh, H. C. (2021). Virtual reality in problem-based learning contexts: Effects on the problem-solving performance, vocabulary acquisition and motivation of English language learners. *Journal of Computer Assisted Learning*, *37*(3), 851–860. https://doi.org/10.1111/jcal.12528

Chen, J. C. (2019). Restoring a “newbie” teacher’s 3D virtual teaching trajectory, resilience, and professional development through action research: A Narrative case study. *TESOL Quarterly*, *54*(2), 375–403. https://doi.org/10.1002/tesq.550

Chen, M. R. A., & Hwang, G. J. (2020). Effects of experiencing authentic contexts on English speaking performances, anxiety and motivation of EFL students with different cognitive styles. *Interactive Learning Environments*, 1–21. https://doi.org/10.1080/10494820.2020.1734626

Chien, S. Y., Hwang, G. J., & Jong, M. S. Y. (2020). Effects of peer assessment within the context of spherical video-based virtual reality on EFL students’ English-Speaking performance and learning perceptions. *Computers & Education*, *146*, 103–751. https://doi.org/10.1016/j.compedu.2019.103751

Choi, D. H., Margitay-Becht, A., Estes, J. S., & Dailey-Hebert, A. (2016). *Emerging tools and applications of virtual reality in education.* Information Science Reference.

Creswell, J. W., & Poth, C. N. (2016). *Qualitative inquiry and research design: Choosing among five approaches*. Sage publications.

Dudeney, G., Hockly, N., & Pegrum, M. (2013). *Digital literacies.* Pearson.

Dzekoe, R. (2020). English language education and digital literacy in the 21st century. In P. Vinogradova & J. K. Shin (Eds.), *Contemporary foundations for teaching English as an additional language: Pedagogical approaches and classroom applications* (pp. 217-245).

Early, M., Kendrick, M., & Potts, D. (2015). Multimodality: Out from the margins of English language teaching. *TESOL Quarterly*, *49*(3), 447–460. https://doi.org/10.1002/tesq.246

Ebadi, S., & Ebadijalal, M. (2020). The Effect of Google Expeditions virtual reality on EFL learners’ willingness to communicate and oral proficiency. *Computer Assisted Language Learning*, 1–25. https://doi.org/10.1080/09588221.2020.1854311

Ghanbarzadeh, R., Ghapanchi, A. H., Blumenstein, M., & Talaei-Khoei, A. (2014). A Decade of research on the use of three-dimensional virtual worlds in health care: A Systematic literature review. *Journal of Medical Internet Research*, *16*(2), e47. https://doi.org/10.2196/jmir.3097.

Han, I. (2020). Immersive virtual field trips in education: A Mixed-methods study on elementary students’ presence and perceived learning. *British Journal of Educational Technology*, *51*(2), 420–435. https://doi.org/10.1111/ bjet.12842

Hundley, M., & Holbrook, T. (2013). Set in stone or set in motion?: Multimodal and digital writing with preservice English teachers. *Journal of Adolescent & Adult Literacy*, *56*(6), 500–509. https://doi.org/10.1002/jaal.171

Jenkins, J. (2009). English as a lingua franca: Interpretations and attitudes. *World Englishes*, *28*(2), 200-207.

Jeong, K. O. (2017). Preparing EFL student teachers with new technologies in the Korean context. *Computer Assisted Language Learning*, *30*(6), 488–509. https://doi.org/10.1080/09588221.2017.1321554

Jiang, L., Yu, S., & Zhao, Y. (2021). Teacher engagement with digital multimodal composing in a Chinese tertiary EFL curriculum. *Language Teaching Research*, *25*(4), 613-632. https://doi.org/10.1177/1362168819864975

Jong, M. S. Y., Tsai, C. C., Xie, H., & Wong, F. K. K. (2020). Integrating interactive learner-immersed video-based virtual reality into learning and teaching of physical geography. *British Journal of Educational Technology*, *51*(6), 2064-2079.

Jucker, A. H., Hausendorf, H., Dürscheid, C., Frick, K., Hottiger, C., Kesselheim, W., Linke, A., Meyer, N., & Steger, A. (2018). Doing space in face-to-face interaction and on interactive multimodal platforms. *Journal of Pragmatics*, *134*, 85–101. https://doi.org/10.1016/j.pragma.2018.07.001

Kaplan-Rakowski, R., & Gruber, A. (2019). Low-immersion versus high-immersion virtual reality: Definitions, classification, and examples with a foreign language focus. In *Proceedings of the Innovation in Language Learning International Conference 2019* (pp. 552-555). Filodiritto Publisher.

Keating, E. (2016). The Role of the body and space in digital multimodality. In A. Georgakopoulou & T. Spilioti (Eds.), *The Routledge handbook of language and digital communication* (pp. 259–272). Routledge.

Kozlova, I., & Priven, D. (2015). ESL teacher training in 3D virtual worlds. *Language Learning & Technology*, *19*(1), 83-101.

Kuru-Gönen, S. İ. (2019). A Qualitative study on a situated experience of technology integration: Reflections from pre-service teachers and students. *Computer Assisted Language Learning, 32*(3), 163-189. https://doi.org/10.1080/09588221.2018.1552974

Li, M. (2020). Multimodal pedagogy in TESOL teacher education: Students' perspectives. *System*, *94*, 102–337. https://doi.org/10.1016/j.system.2020.102337

McFaul, H., & FitzGerald, E. (2019). A Realist evaluation of student use of a virtual reality smartphone application in undergraduate legal education. *British Journal of Educational Technology*, *51*(2), 572–589. https://doi.org/10.1111/bjet.12850

McVee, M. B., Bailey, N. M., & Shanahan, L. E. (2008). Teachers and teacher educators learning from new literacies and new technologies. *Teaching Education*, *19*(3), 197–210. https://doi.org/10.1080/10476210802250216

Miller, S. M. (2007). English teacher learning for new times: Digital video composing as multimodal literacy practice. *English Education*, *40*(1), 61–83.

Philippe, S., Souchet, A. D., Lameras, P., Petridis, P., Caporal, J., Coldeboeuf, G., & Duzan, H. (2020). Multimodal teaching, learning and training in virtual reality: A Review and case study. *Virtual Reality & Intelligent Hardware*, *2*(5), 421-442. https://doi.org/10.1016/j.vrih.2020.07.008

Rance-Roney, J. (2010). Jump-starting language and schema for English-language learners: Teacher-composed digital jumpstarts for academic reading. *Journal of Adolescent and Adult Literacy, 53*(5), 386–395. https://doi.org/10.1598/JAAL.53.5.4

Rowsell, J., & Walsh, M. (2011). Rethinking literacy education in new times: Multimodality, multiliteracies, & new literacies. *Brock Education Journal*, *21*(1), 53–62. https://doi.org/10.26522/brocked.v21i1.236

The New London Group. (1996). A Pedagogy of multiliteracies: Designing social futures. *Harvard Educational Review*, *66*(1), 60–93.

Veliz, L., & Hossein, S. (2020). EAL teachers'(un) preparedness to implement classroom practice for multiliteracies pedagogy. In *Multilingual education yearbook 2020* (pp. 63–80). Springer, Cham.

Walsh, M. (2010). Multimodal literacy: What does it mean for classroom practice? *Australian Journal of Language and Literacy, 33*(3), 211–239.

Walshe, N., & Driver, P. (2019). Developing reflective trainee teacher practice with 360-degree video. *Teaching and Teacher Education*, *78*, 97–105. https://doi.org/10.1016/j.tate.2018.11.009

Wu, W. C. V., Manabe, K., Marek, M. W., & Shu, Y. (2021). Enhancing 21st-century competencies via virtual reality digital content creation. *Journal of Research on Technology in Education*, 1-22. https://doi.org/10.1080/15391523.2021.1962455

Yi, Y., & Choi, J. (2015). Teachers' views of multimodal practices in K-12 classrooms: Voices from teachers in the United States. *TESOL Quarterly*, *49*(4), 838–847. https://doi.org/10.1002/tesq.219

Wu, Y.-T., & Wang, L.-J. (2023). Advancing University EFL Students' Argumentative Essay Writing Performance through Knowledge-Building-based Holistic Instruction. *Educational Technology & Society*, *26*(3), 115-128. https://doi.org/10.30191/ETS.202307_26(3).0009

Advancing University EFL Students' Argumentative Essay Writing Performance through Knowledge-Building-based Holistic Instruction

Ying-Tien Wu[1*] and Li-Jen Wang[2]

[1]Graduate Institute of Network Learning Technology, National Central University, Taiwan // [2]Language Teaching and Research Center, National Yang Ming Chiao Tung University, Taiwan // ytwu@cl.ncu.edu.tw // tommywang@nycu.edu.tw

[*]Corresponding author

(Submitted April 13, 2022; Revised September 23, 2022; Accepted October 3, 2022)

ABSTRACT: Previous research has revealed that university students have multiple learning difficulties in argumentative essay writing (AEW). To address this issue, Knowledge building (KB) pedagogy that aims to create holistic learning environments highlighting idea-refinement, learner agency, and collaborative discourse could be promising. Therefore, this study designed and implemented two KB-based holistic AEW instructions integrating KB pedagogy and explicit instruction on argumentative essay structure and writing conventions. A quasi-experimental design explored the effects of the two holistic KB-based AEW instructions on university EFL students' AEW learning. Two classes of university EFL students were assigned to two instruction groups: The Constant agency enhancement (AE) Instruction group (n = 34) and the Progressive opportunistic collaboration (OC) Instruction group (n = 32). The treatments were two different KB-based holistic AEW instructions for 16 weeks. The participant's perception of learning environments was assessed before and after the instructions to examine if the learning environments created by the two instructions were aligned with KB pedagogy. To investigate the effects of the two instructions on students' AEW performance, the students' argumentative essays were evaluated before, in the middle, and after the instruction. It was found that the two KB-based holistic AEW instructions did align with KB pedagogy but provided university EFL students with distinct and unique learning contexts and opportunities. More importantly, this study also revealed that, compared with the Constant AE instruction, the Progressive OC instruction significantly benefited students more in their argumentative essay writing performance in both the structure and the quality of essays ($p < .05$).

Keywords: Knowledge building, Knowledge building pedagogy, Argumentative essay writing, University students, EFL

1. Introduction

1.1. Argumentative essay writing

In the rapidly developed knowledge-based economy, the ability to make good arguments has become more crucial in today's society (Lam et al., 2018; Matos, 2021). For university students, the ability to write compelling arguments is the defining characteristic of a good student at the undergraduate level (Mitchell, 2000). Educators also advocated that argumentation is a crucial component of university students' academic success and conducted related research (e.g., Liu & Stapleton, 2020; Muller Mirza & Perret-Clermont, 2009).

During the past two decades, research on argumentative essay writing at the university level has been growing (e.g., Awada et al., 2020; Barrot & Gabinete, 2021). For a long time, relevant research on the assessment of argumentative essay writing was mainly conducted along with Toulmin's (1958) model in which a sound argument should consist of five critical elements (including claim, data, warrant, backing, and rebuttal). Based on Toulmin's works, some research focused on analyzing the overall quality of argumentative essays (e.g., Nussbaum & Kardash, 2005; Wolfe, 2011), while others focused on assessing the soundness of argument in argumentative essays in terms of acceptability, relevance, and adequacy (e.g., Hughes & Lavery, 2008; Means & Voss, 1996). To provide more insights into learners' argumentative essay writing, some researchers further advocated the importance of integrating the analyses of the structure of argumentative essays into argumentative essay assessment. For example, Erduran et al. (2004) proposed a revised five-level coding scheme by integrating three elements of data, warrant, and backing into one element – grounds to solve the ambiguities in identifying the data, warrant, and backing have often been found in students' writing when using Toulmin's framework. However, Simon (2008) pointed out that the coding scheme above did not define and consider the quality of claims, grounds, and rebuttals in an argumentative essay. To address this issue, Kathpalia and See (2016) further developed a successful rubric for assessing the quality of argumentative essays in terms of claims, grounds, and rebuttals (for the details of the rubric, please refer to the method section). The rubric developed by Kathpalia and

See (2016) could be used to evaluate students' argumentative essays from both the macro aspect (i.e., the structure of argumentative essays) and the micro aspect (i.e., the quality of argumentative essays), and hence could provide a complete picture of an individual learner's argumentative essay writing performance.

Previous research has advocated that argumentative essay writing is complicated and challenging for university students to learn (Lam et al., 2018; Rapanta et al., 2013). With various argumentative essay assessments, relevant studies have revealed learners' challenges in argumentative essay writing, including (1) poor or missing argumentative essay structure when writing argumentative essays (e.g., Hirose, 2003; Liu & Stapleton, 2014; Osborne, 2010); (2) lacking relevant content knowledge for making arguments and writing conventions (e.g., Bacha, 2010; Barrot & Gabinete, 2021; Butler & Britt, 2011; EI-Henawy et al., 2012; Liu & Stapleton, 2020; Rapanta et al., 2013); (3) showing substantial personal favors but ignoring the counterarguments or having difficulties challenging others' stances (e.g., Toplak et al., 2013; Liu & Stapleton, 2020; Osborne et al., 2013). To enhance students' argumentative essay writing performance, implicit instructions have been largely proven helpful in previous studies to solve the first and the second challenges (e.g., Fan & Chen, 2021; Latifi et al., 2021; Prata et al., 2019). Besides, relevant studies with various instructional strategies have been utilized to address the third challenge. Among the studies addressing the third challenge, some obtained rewarding results (e.g., Hsieh, 2017; Wingate, 2012; Wolfe, 2011; Thompson & Wittek, 2016), while others did not (e.g., Putra et al., 2021). It suggests that more research is needed to examine ways to help students overcome difficulties in constructing counterarguments and rebuttals. In recent relevant research, some pioneer studies on essay writing have revealed collaborative discourse in identifying writing ideas (i.e., Chang & Windeatt, 2016) and facilitating learner autonomy (i.e., Hsieh, 2017). These studies have provided some initial evidence for the effectiveness of collaborative discourse in improving the quality of general writing (i.e., Chang & Windeatt, 2016) and topic-oriented writing (i.e., Hsieh, 2017). Similar to general writing and topic-oriented writing, idea generation and learner autonomy are crucial to argumentative essay writing, suggesting that collaborative discourse could be a helpful strategy to facilitate argumentative essay writing. Moreover, it should be noted that university students may have multiple learning challenges in argumentative essay writing. However, relevant studies addressing helping students overcome multiple learning challenges in argumentative essay writing are still rare.

In short, writing a compelling argumentative essay is a crucial skill that university students need. Research on argumentative essay writing has been conducted to help students to write. However, as mentioned earlier, argumentative essay writing is complicated and challenging for university students. Most empirical studies focused on solving one or two challenges simultaneously; no studies addressed this issue by considering multiple learning challenges with new pedagogies. Therefore, innovative pedagogies are needed to facilitate argumentative essay writing.

1.2. Knowledge building pedagogy

One popular or promising way to solve the challenges that the university students are facing is the use of knowledge building pedagogy. To address language learners' various learning needs as they work towards acquiring good language competence, holistic approaches to design instructions have been adopted in relevant studies, and satisfactory findings were revealed in these studies (e.g., Chiu, 2009; Elovskaya et al., 2019; Goh & Burns, 2012; Tomele, 2015). However, relevant research on applying a holistic approach to AEW instruction is still underexplored. Therefore, this study initially attempted to adopt a holistic approach to AEW instruction design. Goh and Burns (2012) proposed a holistic approach to addressing language learners' cognitive, affective (or emotional), and social needs to help learners acquire good language competence. This holistic approach adopts a socio-cognitive perspective, which considers language learning not only a cognitive but also a social process. In accordance with Goh and Burns (2012), the holistic approach in designing AEW instruction in this study also considered learners' cognitive, affective, and social needs. As a result, the current study adopted Knowledge- building (KB) pedagogy that aims to create a learning environment highlighting idea-refinement, learner agency, and collaborative discourse in the community, which may meet university learners' various needs in the AEW learning process. As Scardamalia and Bereiter (2003) proposed, a KB environment is any environment that enhances collaborative efforts to create and continually improve ideas. The learning environment created by KB pedagogy is also a vibe.

The KB pedagogy was proposed by Scardamalia and Bereiter (2003) to meet the challenges and needs of educating knowledge practitioners in the current knowledge society. The KB pedagogy is based on the premise that authentic creative knowledge work (i.e., the practice of knowledge practitioners) can take place in school classrooms (Bereiter & Scardamalia, 2014). Unlike traditional classroom teaching mainly focuses on acquiring knowledge, the KB pedagogy aims to prepare students as knowledge practitioners through authentic creative knowledge work, such as collaborative inquiry or problem-solving. Based on Popper's (1972) epistemology on ideas, the KB pedagogy emphasizes that all *ideas* proposed by students are of value and should be treated as improvable in creative knowledge work (Bereiter, 2002; Popper, 1972). According to KB pedagogy, the purpose of creative knowledge work in school is to advance the state of knowledge in the *classroom community* through progressive and collective discourse. The teacher becomes a guide rather than a director, and allows students to take over a significant portion of the responsibility for their own learning (Scardamalia & Bereiter, 1991). Students are epistemic agencies that actively engage in negotiation and dialogue to fit personal ideas with others (Scardamalia & Bereiter, 2006). The definition of an agency in KB is a learner who is expected to take a more active role in knowledge advancement or idea refinement. (Scardamalia & Bereiter, 2003). To support the implementation of KB pedagogy, an online platform, Knowledge Forum (KF), was designed and developed. Learners could create, refine, and integrate ideas by writing notes on KF. The KF platform was developed to facilitate learners to contribute ideas, rise above their pre-existing understandings, and improve their community knowledge (Hong & Scardamalia, 2015; Sun et al., 2010; Wu & Wang, 2016). Similar to other collaborative learning software or online learning platforms, the KF also provides a set of scaffolds (similar to openers utilized in other collaborative or argumentation learning software) to support students in developing the content of notes. These scaffolds could help students clarify and organize their ideas or arguments when writing notes for different aspects of knowledge building processes (Scardamalia, 2004). A set of scaffolds has also been designed and used for supporting collaborative argumentation learning in previous research (e.g., Wu et al., 2017).

1.3. KB-based holistic argumentative essay writing instructions

To help students overcome learning challenges in argumentative essay writing, two holistic instructions integrated KB pedagogy and explicit instruction on argumentative essay structure and writing conventions were designed, implemented, and examined in this study. In relevant studies, researchers have developed two different types of KB-based instructions, agency-enhancement KB-based instruction (e.g., Hong et al., 2020; Zhang et al., 2009) and opportunistic-collaboration KB-based instruction (e.g., Hung & Hong, 2017; Sawyer, 2007; Zhang et al., 2009). In KB pedagogy, a classroom community may be broadly or narrowly defined (Scardamalia & Bereiter, 2006). Agency-enhancement KB-based instruction is usually implemented with fixed and small student groupings, and each student group is viewed as a classroom community. It mainly focuses on promoting students' agency to engage in idea-centered learning actively and has been advocated as an effective method of transferring more responsibility to students (e.g., Zhang et al., 2009). As it is easier to be implemented in traditional classroom settings, it is the most common KB-based instruction in previous studies. Moreover, to provide students with a more authentic KB experience (i.e., the practice of knowledge practitioners), some relevant studies also implemented opportunistic-collaboration KB-based instruction (e.g., Siqin et al., 2015). It emphasizes working with ideas, assuming agency, and fostering a highly culture-related community-wide collaboration. In opportunistic-collaboration KB-based instruction, the whole class is viewed as one community, and all students are invited and seen as a part of the classroom community (i.e., all students are in one big group and have the same responsibility to collaborate with others). Based on the two KB-based instructions above, two holistic instructions on argumentative essay writing (AEW), Constant Agency-enhancement Instruction (Constant AE Instruction) and Progressive Opportunistic-collaboration Instruction (Progressive OC Instruction) were designed in this study. Typically, Taiwanese students receive test-oriented and teacher-centered instructions in high schools. As a result, first-year university students often lack agency and become less active learners (Hsu, 2015). Besides, most of them had the experience of discussing in groups rather than in a big community of students in the classroom. Based on students' prior learning experience above, an agency-enhancement KB-based instruction focusing on improving student agency could be suitable to be implemented in the freshman AEW courses. Therefore, the first holistic AEW instruction (called Constant AE Instruction) was designed by integrating agency-enhancement KB-based instruction with fixed-small grouping and explicit instruction on argumentative essay structure and writing conventions. Moreover, previous research has advocated that opportunistic collaboration instruction could provide more authentic knowledge building experiences (e.g., Siqin et al., 2015; Zhang et al., 2009). In Siqin et al. (2015), a progressive KB-based instruction in which fixed-group collaboration was combined with opportunistic collaboration was designed for a KB-based undergraduate course. They divided the 16-week course into two equal phases. During the first phase (8 weeks), students were randomly assigned to groups of five or six and discussed their ideas with their group members. During the second phase (8 weeks), all students formed one big group but were allowed to form various collaborative groups at their discretion. As most first-year university students in Taiwan do not have experience in community-based discourse, this study designed the second holistic AEW instruction (called Progressive OC Instruction) to enculturate students into a knowledge building paradigm gradually. This holistic AEW instruction integrated the explicit instruction on argumentative essay structure, writing conventions, and progressive KB-based instruction adapted from Siqin et al. (2015).

In this study, the two KB-based holistic AEW instructions were implemented and evaluated in two university AEW courses, respectively. Besides, in contrast with learning that is focused on knowledge acquisition in traditional classrooms, a knowledge building environment encourages learners to produce diverse ideas and develop, refine or elaborate the ideas through progressive discourse. Therefore, to examine if the learning environments created by the two KB-based holistic AEW instructions developed in this study were aligned with KB pedagogy, the student's perceptions of their conventional EFL learning environments were assessed before and after the two holistic AEW instructions. In sum, two major research questions were proposed in this study:

- What are the effects of the two different KB-based holistic AEW instructions on university EFL students' perceptions of learning environments?
- What are the effects of the two different KB-based holistic AEW instructions on university EFL students' argumentative essay performance?

2. Methods

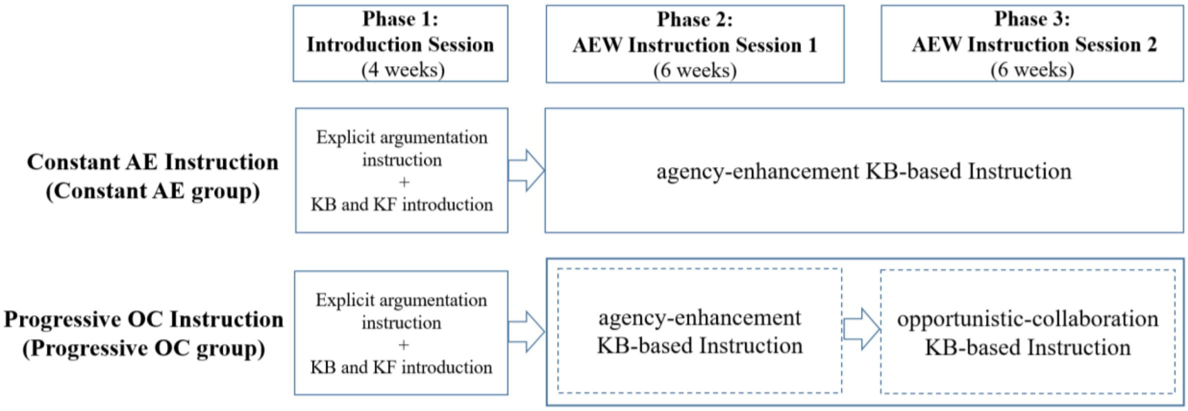
2.1. Participants

The participants were two classes of first-year non-English majors (*n* = 66) at a university in northern Taiwan. The participants' English proficiency levels were at CEFR B1 to B2, meaning they understand the main ideas when reading a complex text, as well as contemporary literary prose, articles, and reports. Also, they could write clear, detailed texts on subjects related to their interests or expertise. They were voluntarily enrolled in a Freshman English focusing on argumentative essay writing taught by the same instructor who has had the experience of adopting KB-based instructions at the university level for over five years. The instructor played the role of guidance and gave instructions and feedback to the students in the study. The classes were face-to-face and met once a week for 2 hours. The participants were required to attend the in-class and after-class learning tasks. They were native speakers of Mandarin Chinese and had studied English for approximately six years or above. The study's participants had never taken courses focusing on argumentative essay writing. However, they had been taught to write a variety of genres, such as narrative, description, and exposition, in senior high school. The participants have been randomly placed into the two classes in this study. Because they had only experience with general English writing or topic-oriented writing in senior high school, their English argumentative essay writing abilities were reasonably poor, and they had not received any KB instructions before the courses.

2.2. Research design

This study adopted a quasi-experimental research method. The treatment of this study was two different KB-based holistic AEW instructions, and the two classes of students were assigned to two different instruction groups: The Constant AE group (*n* = 34) and the Progressive OC group (*n* = 32). Both the two different KB-based holistic AEW instructions included an introduction session and two AEW instruction sessions (please see Figure 1). The power shortage issue in Taiwan and the use of genetic-modified food were selected as the topics for the two AEW instruction sessions, respectively. As university students in Taiwan have studied the introduction of various energy power and genetic modification in high school science class, the participants in this study had a basic understanding of various energy power and genetic modification. Besides, the power shortage issue in Taiwan and the use of genetically-modified food are daily issues relevant to them. Therefore, all the participants had basic prior knowledge and shared an interest in the two topics.

Figure 1. KB-based holistic AEW instructions



Before this study, the students were asked to write an individual argumentative essay regarding The power shortage issue in Taiwan as the pre-test. Then, after the first AEW instruction session, they were asked to write an individual argumentative essay regarding the same topic again as the middle test. Finally, in the post-test, each student was asked to write an argumentative essay regarding the use of genetic-modified food as the post-test after the second AEW instruction session. The three argumentative essays were evaluated to examine the effects of the two different KB-based holistic AEW instructions on students' argumentative writing essay writing performance. Besides, to examine if the learning environments created by the two KB-based holistic AEW instructions developed in this study were aligned with KB pedagogy, the two groups of students' perceptions of learning environments were assessed before and after the KB-based holistic AEW instructions.

2.3. KB-based holistic AEW instructions

As shown in Figure 1, the two holistic AEW instructions included three phases. For the two holistic AEW instructions, in the first phase (i.e., Introduction Session), the instructor gave lectures about the elements of argumentation and the structure of the argumentative essay. The knowledge building theory and Knowledge Forum were briefly introduced to the students. After Phase 1, two KB-based AEW Instruction sessions were implemented in Phase 2 and Phase 3. For the Constant AE group, agency-enhancement KB-based instruction was constantly implemented in the two AEW instruction sessions. In contrast, in the Progressive OC group, the Progressive OC instruction was implemented to progressively enculturate students with discourse-based idea refinement and knowledge advancement within a KB community. The students in the Progressive OC group received an agency-enhancement KB-based instruction in the first AEW Instruction session (i.e., Phase 2). Then the KB-based instruction was transformed into an opportunistic-collaboration KB-based Instruction in the next session (i.e., Phase 3). At the end of Phase 2 and phase 3, each student in the two groups was required to finish a five-paragraph argumentative essay on the power shortage issue in Taiwan and the use of genetic-modified food, respectively (at least 500 words or above).

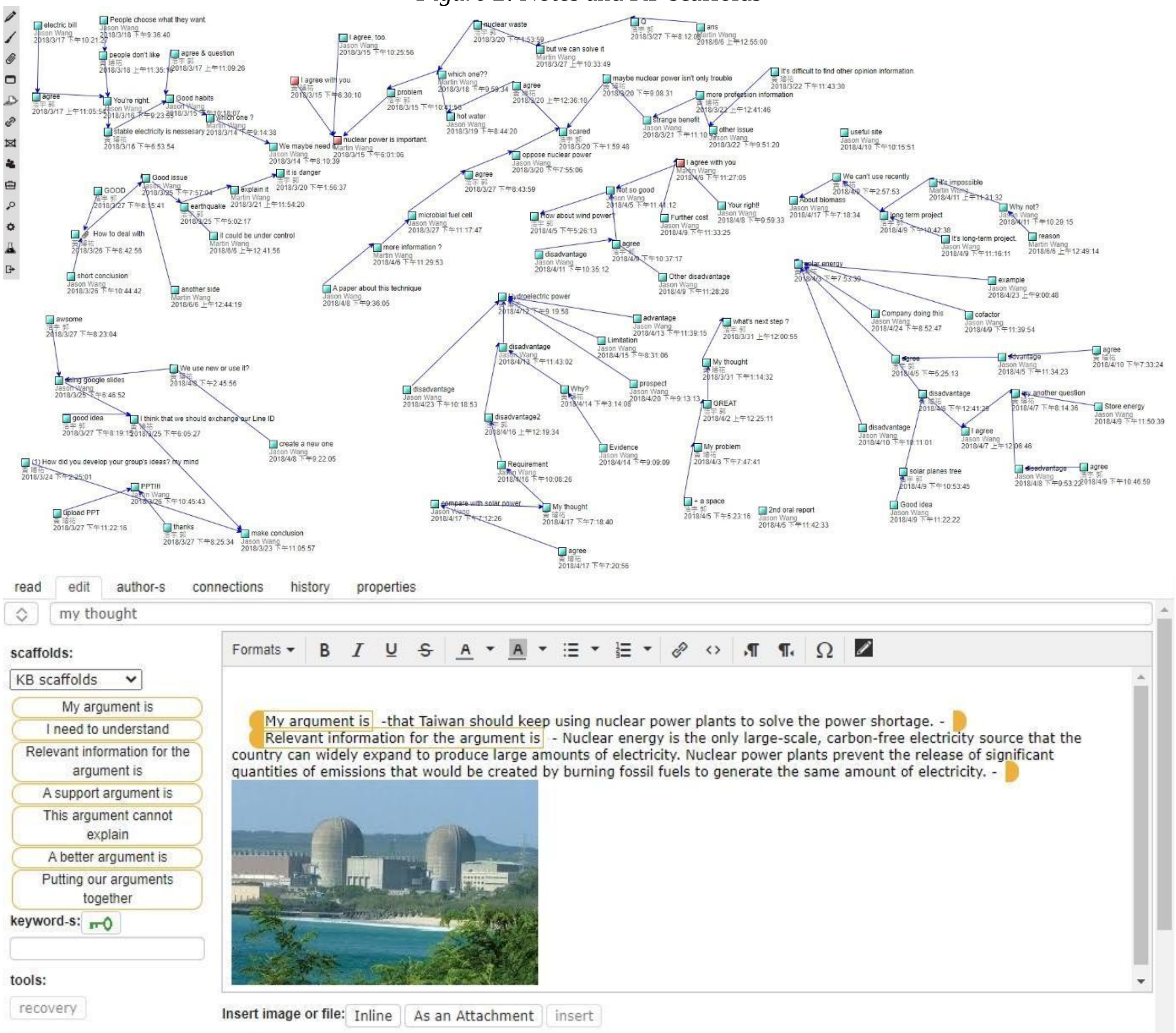
Both agency-enhancement KB-based instruction and opportunistic-collaboration KB-based instruction implemented in the AEW instruction session were designed based on the constructive alignment framework proposed by Biggs (2003). According to Biggs' (2003) framework, when designing instruction, it should start with the desired learning outcomes that we intend students to learn, and the learning activities and assessment have to be aligned with the desired outcomes. Thus, learners could construct meaning from what they do to learn. Based on Biggs' (2003) framework, the desired learning outcome in the AEW instruction session in this study is a satisfactory individual argumentative essay regarding a specific issue. To this end, the learners were asked to obtain relevant information with collaborative inquiry activity, generate personal arguments and learn arguments from alternative perspectives with idea-centered collaborative argumentation, integrate arguments from various perspectives with group reports and reflection, and eventually write individual argumentative essays as their learning outcomes. In this study, both agency-enhancement KB-based instruction and opportunistic-collaboration KB-based instruction implemented in the AEW instruction session involved four major learning activities in a row: (1) *Collaborative inquiry activity*: To improve students' relevant content knowledge regarding the topic for making arguments and writing conventions, students were guided to search for relevant information on the internet and shared what they learned about the topic on KF within the classroom community in and out of class. (2) *Idea-centered collaborative argumentation*: To improve students' ability to generate arguments and allow them to argue from different perspectives, students had to produce their ideas regarding the issue and generate relevant arguments with their prior knowledge and the relevant information found. They were also asked to generate and share and construct various arguments to respond to others' arguments collaboratively in and out of class. They were guided to propose their evidence-based arguments actively on KF, which provides various argumentation prompts as scaffolds for argument generation. Then, through collaborative discourse for integrating and advancing community knowledge, they could clarify and refine the ideas proposed by community members. During the process, arguments from different perspectives could be proposed, and strong personal favors might be diminished. (3) *Group report and reflection*: After the idea-centered collaborative argumentation activities, students were required to give oral reflective reports regarding their knowledge building and collaborative argumentation practice every week. Based on that, they were also asked to reflect on how to generate the arguments and organize the structure of their argumentative essays. (4) *Individual essay writing*: Finally, each student had to write an individual argumentative essay based on the information and argumentation regarding the topics. They were required to write a five-paragraph argumentative essay with at least 500 words.

Similar to Zhang et al. (2009), the agency-enhancement KB-based instruction in the AEW instruction session was implemented with fixed and small student groups (3 to 4 students), and each student group was viewed as a classroom community. It highlighted to transfer more responsibility to students during the learning activities. As a result, the instructor's guide for the learning activities during the instruction session mainly focused on promoting students' agency to engage in idea-centered learning activities. The opportunistic-collaboration KB-based instruction in the AEW instruction session aimed to mainly focus on advancing the state of knowledge in the class while situating it within the broader societal effort to build knowledge. The whole class is viewed as one community, and all students are invited and seen as a part of the classroom community. To this end, the instructor's guidance during the instruction session focused on working with ideas, assuming agency, and fostering a highly culture-related community-wide collaboration.

2.4. Knowledge Forum (KF) as an online collaboration platform

This study adopted Knowledge Forum (KF) as an online collaboration platform. KF was designed to support idea work and move it to higher levels. As shown in Figure 2, the seven scripted scaffolds embedded in KF included: My argument is, I need to understand, Relevant information for the argument is, A supportive argument is, and This argument cannot explain; A better argument is, and Putting our arguments together. First, students made the notes themselves, and other group members built on the notes with scaffolding annotations. Then, students needed to explain their purposes for responding using the scripted scaffolds embedded in KF. During KB-based holistic AEW instructions, students could have opportunities to share and further enhance their content knowledge by taking notes, and their tendency to show personal favoritism could be reduced during community discourse. Also, the scaffolds provided by KF could make students propose arguments purposefully when making a note on KF. It could help improve the quality of students' arguments.

Figure 2. Notes and KF scaffolds



2.5. Instruments

2.5.1. Knowledge Building Environment Scale (KBES)

To evaluate the participants' perception of learning environments created by the two KB-based holistic AEW instructions, the Knowledge Building Environment Scale (KBES) was used. The KBES was developed by Lin et al. (2014) to evaluate university students' perception of learning environments from the perspectives of KB pedagogy. The KBES was a four-point Likert scale, and it consists of three subscales echoing the KB pedagogy, including working with ideas, assuming agency, and fostering community. There is a total of 24 items in the KBES. Through a series of confirmatory factor analyses (CFA), the validity of KBES was confirmed by Lin et al. (2014). Besides, they also reported the Cronbach's alpha coefficients for the three subscales of KBES as 0.85, 0.91, and 0.94, respectively, revealing that the KBES was deemed to be sufficiently reliable for assessing students' perception of learning environments from the perspectives of KB pedagogy. Thus, the KBES was an effective tool for measuring perceptions of a knowledge building environment among students and was therefore employed in this study. In this study, the Cronbach's alpha coefficients for the three subscales of KBES were 0.82, 0.84, and 0.90, respectively, and the overall Cronbach's alpha coefficient for the 24 items was 0.91.

2.5.2. Coding scheme of argumentative essay structure and quality

With the coding scheme in Kathpalia and See (2016), the three argumentative essays completed in the pre-test, middle-test, and post-test were evaluated from the two aspects, including the structure (Macro view) and the quality (Micro view) in this study. First, for the Macro view aspect, the three-level coding scheme in Kathpalia and See (2016) was used for evaluating the structure of argumentative essays, namely, Lower level: simple claim or grounds only; Intermediate level: claim with valid grounds; Higher level: rebuttal with a clear claim but partial evidence in the form of a warrant, rebuttal with a clear claim and grounds, or extended argument with a claim supported by grounds with more than one rebuttal. For the micro view aspect, the participants' argumentative essay quality was assessed with the rubric developed in Kathpalia and See (2016), as shown in Table 1.

Table 1. Argumentation quality of elements coding scheme

Rubrics		Description	Score
Claims	Weak claim	One that fails to address the proposition mentioned in the argumentative essay	1
	Strong claim	One that addresses the proposition mentioned in the argumentative essay	2
Grounds	No evidence	Just personal opinions	0
	Faulty evidence	Weak evidence refers to faulty evidence	1
	Personal only	Intermediate evidence refers to personal evidence	2
	Attributive only	Intermediate evidence refers to attributive evidence	3
	Attributive & personal	Strong evidence refers to an attributive or a combination of attributive and personal evidence	4
Rebuttals	No rebuttal	no counter-view	0
	Weak rebuttals	Only contains a counter-view without a rebuttal	1
	Strong rebuttals	Contains a counter-argument and rebuttal	2

The authors and one of their colleagues (another EFL lecturer at one of the authors' university) coded and assessed the data independently regarding the argumentation structure and quality of elements based on the two coding schemes mentioned above. Their coding results were compared, and the inter-coder reliability in each coding category was higher than 0.87, showing a high consistency in the researchers' coding of the data set. Then, all the differences were resolved through discussions.

3. Results

3.1. The effects of holistic KB-based AEW instructions on students' perceptions of learning environments

As shown in Table 2, in the Constant AE group, significant differences between the students' responses in all the three subscales of the KBES in the pretest and the post-test were revealed ($p < .001$). Similar results were also found in the Progressive OC group ($p < .001$). It indicated that, compared with the instruction in the participants' previous courses, both the Constant AE instruction and Progressive OC instruction did provide relatively more opportunity for students to work with ideas and engage in exploring the issues of argumentative essays actively to work collaboratively as a community. The two KB-based holistic AEW instructions in this study did align with knowledge building pedagogy.

Table 2. The students' perception of learning environments before and after the KB-based AEW instructions

		Pretest		Post-test		t value
		Mean	*SD*	Mean	*SD*	
Working with ideas	AE group (n = 34)	2.34	0.29	3.25	0.40	11.98^{***}
	OC group (n = 32)	2.33	0.26	3.47	0.23	19.10^{***}
Assuming agency	AE group (n = 34)	2.70	0.35	3.15	0.43	5.05^{***}
	OC group (n = 32)	2.81	0.31	3.30	0.30	7.86^{***}
Fostering community	AE group (n = 34)	2.66	0.32	3.12	0.37	4.78^{***}
	OC group (n = 32)	2.73	0.31	3.42	0.17	9.60^{***}

Note. $^{***}p < .001$.

Also, a one-way ANCOVA (the students' responses on the KBES in the pre-test were used as the covariate) was conducted to compare the effects of the two KB-based holistic AEW instructions on the students' experiences of knowledge building. As shown in Table 3, significant differences were found in the two subscales, working with ideas and fostering community ($p < .01$), suggesting that compared with the Constant AE instruction, the Progressive OC instruction provided students more opportunities to work with ideas and have broader community collaboration among students. However, there was no significant difference between the two groups in Assuming agency, suggesting that the two holistic instructions could equally help students become more active learners.

Table 3. The results of ANCOVA in the students' perception of learning environments

		Mean (adjusted)	Standard error	F-value
Working with ideas	AE groups (n = 34)	3.26	0.53	7.1^{**}
	OC groups (n = 32)	3.46	0.54	
Assuming agency	AE groups (n = 34)	3.15	0.42	2.29
	OC groups (n = 32)	3.30	0.29	
Fostering community	AE groups (n = 34)	3.12	0.37	15.41^{***}
	OC groups (n = 32)	3.42	0.16	

Note. $^{**}p < .01$; $^{***}p < .001$.

3.2. The effects of holistic KB-based AEW instructions on students' argumentative essay performance

3.2.1. Argumentation structure level

In this study, as Kathpalia and See (2016), the students' argumentative essay structure level was categorized into three levels: lower, intermediate, and higher. Three chi-square tests were conducted in the pretest, middle-test, and post-test to examine whether the students in the two instruction groups have the same proportions at the three argumentative essay structure levels.

Table 4. Group comparisons on the students' writing skills in organizing the structure of argumentative essay

		Lower (n, %)	Intermediate (n, %)	Higher (n, %)
Pretest	Constant AE group	17 (50%)	15 (44%)	2 (6%)
	Progressive OC group	16 (50%)	13 (41%)	3 (9%)
	X^2 value		0.31 (n.s.)	
Middle-test	Constant AE group	2 (6%)	4 (12%)	28 (82%)
	Progressive OC group	1 (3%)	4 (12%)	27 (85%)
	X^2 value		0.29 (n.s.)	
Post-test	Constant AE group	0 (0%)	7 (21%)	27 (79%)
	Progressive OC group	0 (0%)	1 (3%)	31 (97%)
	X^2 value		4.72^{*}	

Note. n.s.: non-significant; $^{*}p < .05$.

As shown in Table 4, no significant difference was found in the pre-test and middle-test ($p > .05$). It indicated that the two groups of students did not have significant differences in the argumentation structure levels of their essays before the conduct of two instructions (i.e., phase 1). Also, they have no significant difference in the argumentation structure levels of their essays after the AEW instruction session 1 (i.e., phase 2). However, significant differences were found in the post-test (p < .05). It should be noticed that almost all of the students who received Progressive OC instruction (97%) achieved a higher level, while only about three-quarters of the students who received Constant AE instruction (79%) achieved a higher level and about a quarter (21%) of them achieved an intermediate level. It suggested that compared with the Constant AE instruction, the Progressive OC instruction could benefit EFL students more in improving their performance in arranging the structures of argumentative essays.

3.2.2. Argumentation quality

The students' argumentative essay quality was assessed in three elements: claims, grounds, and rebuttals. According to the coding scheme used in this study, students' quality of claims revealed in their argumentative essays was classified into two levels (i.e., weak and strong). The quality of grounds was divided into four levels (i.e., faulty evidence, personal only, attribute only, attributive and personal). The quality of rebuttals was classified into two levels (i.e., weak and strong). Three chi-square tests were conducted to examine whether the students in the two instruction groups had the same proportions of their quality of claims as each other in the pre-test, middle-test, and post-test. As revealed in Table 5, there were no significant differences between the two groups on the quality of claims in the pre-test (χ^2 = 0.96, p > .05), the middle-test (χ^2 = 0.29, p > .05), and the post-test (χ^2 = 1.61, p > .05). It should be noticed that very high proportions of students in both groups (82% for the Constant AE group and 91% for the Progressive OC group) could generate claims of higher quality (i.e., strong claims). After instruction, the high proportions of strong claims remained in both student groups.

Table 5. Group comparisons on the quality of claims in the three argumentative essays

		Weak claim	Strong claim	X^2
Pre-test	Constant AE group	6(18%)	28(82%)	0.96 (n.s.)
	Progressive OC group	3(9%)	29(91%)	
Middle-test	Constant AE group	2(6%)	32(94%)	0.29 (n.s.)
	Progressive OC group	3(%)	29(91%)	
Post-test	Constant AE group	7(21%)	27(79%)	1.61 (n.s.)
	Progressive OC group	3(%)	29(91%)	

Note. n.s.: non-significant.

Table 6 revealed that no significant differences were found between the two groups on the quality of grounds in both the pre-test (χ^2 = 4.56, p > .05) and the middle-test (χ^2 = 3.47, p > .05). It indicated that the two groups of students did not have significant differences in the quality of grounds in their essays before the conduct of two instructions (i.e., phase 1). Also, they have no significant differences in the quality of grounds after the AEW instruction session 1 (i.e., phase 2). However, significant differences were found in the post-test (χ^2 = 17.06, p < .05). It is worth noting that over half of the students who received Progressive OC instruction (60%) cited both attribute and personal grounds, which is a more persuasive approach for supporting a claim. In contrast, only one-third of the students who received Constant AE instruction (30 %) adopted this approach. Therefore, it seems that, compared with the Constant AE instruction, the Progressive OC instruction could be more capable of improving university students' quality of grounds in their argumentative essays.

Table 6. Group comparisons on the quality of grounds in the three argumentative essays

		Faulty evidence	Personal only	Attribute only	Attribute and personal	X^2
Pre-test	Constant AE group	8(23%)	22(65%)	4(12%)	0(0%)	4.56 (n.s.)
	Progressive OC group	6(19%)	26(81%)	0(0%)	0(0%)	
Middle-test	Constant AE group	1(3%)	12(35%)	8(24%)	13(38%)	3.47 (n.s.)
	Progressive OC group	0(0%)	6(19%)	10(31%)	16(50%)	
Post-test	Constant AE group	1(2%)	17(50%)	6(18%)	10(30%)	17.06*
	Progressive OC group	0(0%)	2(6%)	11(34%)	19(60%)	

Note. n.s.: non-significant; *p < .05.

As shown in Table 7, no significant differences were found between the two groups on the quality of rebuttals in both the pre-test (χ^2 = 0.29, p < .05) and the middle-test (χ^2 = 0.23, p < .05). However, similar to their quality of grounds, significant differences between the two groups of students were only found in the post-test (χ^2 = 4.18, p < .05). Most students who received Progressive OC instruction (91%) discovered how to provide counterarguments and rebuttals, which is a more compelling way to support their positions. Comparatively, three-fourths of the students receiving Constant AE instruction (76 %) adopted this approach and learned to use rebuttals to enhance their essays. It seems that the Progressive OC instruction could benefit university students more in improving the quality of rebuttals in their argumentative essays.

Table 7. Group comparisons on the quality of rebuttals in the three argumentative essays

		Weak rebuttals	Strong rebuttals	X^2
Pre-test	Constant AE group	32(94%)	2(6%)	0.29 (n.s.)
	Progressive OC group	29(91%)	3(9%)	
Middle-test	Constant AE group	8(24%)	26(76%)	0.23 (n.s.)
	Progressive OC group	6(19%)	26(81%)	
Post-test	Constant AE group	8(24%)	26(76%)	4.18^*
	Progressive OC group	3(9%)	29(91%)	

Note. n.s.: non-significant; $^*p < .05$.

4. Discussion

In this study, two different KB-based holistic AEW instructions were developed, and the effects of the two AEW instructions on university EFL students' perceptions of learning environments were investigated. KB pedagogy was applied in the two KB-based holistic AEW instructions to help students overcome AEW learning challenges with collaborative discourse. Both instructions were confirmed to align with KB pedagogy and be capable of cultivating learning environments that shared major features of KB pedagogy for AEW learners. However, this study also revealed major significant differences between the learning environments created by the two instructions. In particular, the Progressive OC instruction provided students with a better AEW learning environment than the Constant AE instruction. More opportunities were provided for students to have community collaboration and collaborative discourse. It suggests that the Progressive OC instruction design that gradually enculturates students into a knowledge building paradigm could be implementable and effective for EFL university learners in learning argumentative essay writing.

Moreover, this study revealed the effectiveness of the two AEW instructions on university EFL students' argumentative essay performance. In this study, an integrated coding scheme with both macro and micro aspects was used to assess students' argumentative essays. The major findings regarding the effects of the two AEW instructions on university EFL students' argumentative essay performance are discussed from the two aspects, respectively. As for the macro aspect, it was found that the two KB-based holistic AEW instructions could improve students' writing skills in organizing the structures of their argumentative essays. Matos (2021) claimed that engagement in this collaborative writing process could offer a promising path to enhancing argumentative essay structure. Also, Resnick et al. (2015) advocated that the collaborative writing process could allow individual learners to advance their thinking and writing by deeper engagement, and collaborative writing could build a bridge between peer discourse and personal writing, providing rich cognitive context for developing argumentation skills. In this study, during the KB-based holistic AEW instructions with KF, students had undergone a collaborative writing process in which they were required to discuss and generate their arguments and evidence with their peers before writing their argumentative essays. In addition, they needed to propose their claims with concrete arguments, provide both subjective and objective grounds, and offer counterarguments and rebuttals. Therefore, similar to the collaborative writing process mentioned in Matos (2021) and Resnick et al. (2015), the online collaborative discourse process on the KF in this study seemed to provide the students with opportunities to identify the crucial elements of argumentative essay structure (i.e., claims, grounds, and rebuttals). Consequently, the students' macro views regarding argumentative essays could be shaped and developed, contributing to the improvement of their individual argumentative essay writing structure. Regarding the micro aspect, this study found that the KB-based holistic AEW instructions in the Constant AE and Progressive OC groups could enhance students' argumentative essay quality. Moreover, it should be noticed that in each session of both the AEW instructions in this study adopted four consistent learning activities step by step. As Biggs (2003) suggested, "The students are 'entrapped' in this web of consistency, optimizing the likelihood that they will engage in appropriate learning activities, but paradoxically frees students to conceal their own learning" (p. 26). The results have proved that the four-step progressive design for the explicit AEW instruction sessions benefited students' performance in AEW.

This study also revealed that the Progressive OC instruction progressively engaging the students in group-based collaboration to community-based collaboration could benefit students more than the Constant AE instruction. As aforementioned, first-year university students in Taiwan typically receive test-oriented and teacher-centered instructions in high schools. Moreover, most of them have experience in fixed small group discussions but do not have experience in community-based discourse. Undoubtedly, it would be a better way to progressively design learning activities based on students' prior experiences and knowledge. Therefore, based on the participants prior learning experience in high school, an explicit AEW instruction with a fixed-small grouping was implemented in the first session of the Progressive OC Instruction to promote students' agency to engage in idea-centered learning activities. After the first session of the Progressive OC Instruction, students' learning experience in idea-centered and group-based collaboration was advanced, which could serve as an essential foundation of an idea-centered and community-wide collaboration. Furthermore, Putra et al. (2021) have confirmed that opportunistic collaboration writing could facilitate students' grounds and rebuttals in their argumentative essays. Moreover, Kathpalia and See (2016) also advocated that students with more freedom and responsibility could engage in different ideas and expand the diversity of their ideas in their argumentative essay writing. The Progressive OC instruction allowed the students more freedom and responsibility to propose, refine and integrate ideas from diverse perspectives when generating arguments on KF. Hence, their understanding regarding grounds and rebuttals could be gradually better shaped and developed during the Progressive OC instruction, and then it could serve as a superior foundation for proposing more insightful grounds and rebuttals in their argumentative essay writing.

5. Conclusion

As one of the initial attempts, this study applied KB pedagogy in designing KB-based holistic AEW instruction. The findings derived from the current study provide some important implications for educational practice. Collaboration activities, such as collaborative writing in Kathpalia and See (2016) or online collaborative argumentation activities in this study, could improve the quality of students' argumentative essays. The findings of this study highlight the feasibility of KB pedagogy in argumentative essay writing instruction. EFL teachers could design AEW learning activities emphasizing the three core dimensions of KB pedagogy. Learning activities should provide opportunities for students to work closely with ideas, actively explore the issues of argumentative essays, and aggressively form a community.

6. Limitations

One may be interested in comparing the effectiveness of KB-based holistic AEW instruction with conventional AEW instruction. However, this study is limited to provide insights into the aforementioned issue. This study only investigated the effectiveness of KB-based holistic AEW instruction by comparing the effects of two different KB-based holistic AEW instructions. The effectiveness of the KB-based holistic AEW instructions was not compared with conventional AEW instruction in this study. To address this issue, a follow-up quasi-experimental study could be conducted with adding a conventional AEW instruction group.

7. Suggestions for future research

This study provides some important directions for future research. First, future practical work could focus on developing other specific holistic AEW instructional strategies or modifying this study's holistic AEW instruction design based on KB's twelve principles. One may be also curious about how students engage in Progressive OC instruction and transform their artifacts from their collaborative argumentation on the KF into their argumentative essay writing. To address this issue, future research could be conducted. Also, how students collaborate during KB-based holistic AEW learning and their individual learning process are still underexplored.

Acknowledgment

The funding of the research work was supported by National Science and Technology Council, Taiwan, under grant numbers MOST 107-2628-H-008 -003 -MY4 and MOST 111-2410-H-008 -004 -MY3.

References

Awada, G., Burston, J., & Ghannage, R. (2020). Effect of student team achievement division through WebQuest on EFL students' argumentative writing skills and their instructors' perceptions. *Computer Assisted Language Learning, 33*(3), 275-300. https://doi.org/10.1080/09588221.2018.1558254

Bacha, N. (2010). Teaching the academic argument in a university EFL environment. *Journal of English for Academic Purposes, 93*, 229–241. https://doi.org/10.1016/j.jeap.2010.05.001

Barrot, J., & Gabinete, M. K. (2021). Complexity, accuracy, and fluency in the argumentative writing of ESL and EFL learners. *International Review of Applied Linguistics in Language Teaching, 59*(2), 209-232. https://doi.org/10.1515/iral-2017-0012

Bereiter, C. (2002). *Education and mind in the knowledge age.* Lawrence Erlbaum Associates. https://doi.org/10.4324/9781410612182

Bereiter, C., & Scardamalia, M. (2014). Knowledge building and knowledge creation: One concept, two hills to climb. In S. C. Tan, H. J. So, & J. Yeo (Eds.), *Knowledge creation in education* (pp. 35-52). Springer.

Biggs, J. B. (2003). *Teaching for quality learning at university* (2nd ed.). Open University Press/Society for Research into Higher Education.

Butler, J. A., & Britt, M. A (2011). Investigating instruction for improving revision of argumentative essays. *Written Communication, 28*(1),70-96.

Chang, H., & Windeatt, S. (2016). Developing collaborative learning practices in an online language course. *Computer Assisted Language Learning, 29*(8), 1271-1286. http://doi.org/10.1080/09588221.2016.1274331

Chiu, Y. J. (2009). Exploring pedagogical strategies to integrate a holistic approach to education into ESL classes. *Asia Pacific Education Review, 10*(2), 261-270.

EI-Henawy, W. M., Dadour, E.-S. M., Salem, M. M., & EI-Bassuony, J. M. (2012). The Effectiveness of using self-regulation strategies on developing argumentative writing of EFL prospective teachers. *Journal of the Egyptian Association for Reading and Knowledge, 27*(1), 1-28.

Elovskaya, S., Stanchuliak, T., & Karandeeva, L. (2019). The Holistic approach to teaching English as a foreign language. In *INTED2019 Proceedings* (pp. 1253-1258). IATED. http://doi.org/10.21125/inted.2019.0402

Erduran, S., Simon, S., & Osbrne, J. (2004). TAPping into argumentation: developments in the application of Toulmin's argument pattern for studying science discourse. *Science Education, 88*(6), 915-933. http://doi.org/10.1002/sce.20012

Fan, C. Y., & Chen, G. D. (2021). A Scaffolding tool to assist learners in argumentative writing. *Computer Assisted Language Learning, 34*(1-2), 159-183. http://doi.org/10.1080/09588221.2019.1660685

Goh, C. C. M., & Burns, A. (2012). *Teaching speaking: A Holistic approach.* Cambridge University Press.

Hirose, K. (2003). Comparing L1 and L2 organizational patterns in the argumentative writing of Japanese EFL students. *Journal of Second Language Writing, 12*, 181-209. https://doi.org/10.1016/S1060-3743(03)00015-8

Hong, H.-Y., & Scardamalia, M. (2015). Community knowledge assessment in a knowledge building environment. *Computer & Education, 71*(2014), 279-288. https://doi.org/10.1016/j.compedu.2013.09.009

Hong, H.-Y., Ma, -L., Lin, P.-Y., & Lee, Y.-H. (2020). Advancing third graders' reading comprehension through collaborative knowledge building: A Comparative study in Taiwan. *Computers & Education, 157*(11), 1-11. https://doi.org/10.1016/j.compedu.2020.103962

Hsieh, Y. C. (2017). A Case study of the dynamics of scaffolding among ESL learners and online resources in collaborative learning. *Computer Assisted Language Learning, 30*(1-2), 115-132. http://doi.org/10.1080/09588221.2016.1273245

Hsu, W.-H. (2015). Transitioning to a communication-oriented pedagogy: Taiwanese university freshmen's views on class participation. *System, 49*, 61-72. https://doi.org/10.1016/j.system.2014.12.002

Hughes, W., & Lavery, J. (2008). *Critical thinking: An Introduction to the basic skills* (5th ed.). Broadview Press.

Hung, G. T., & Hong, H.-Y. (2017). Fostering innovation-oriented cultural views among college students through knowledge-building pedagogy. *Learning: Research and Practice, 3*(2), 168-182. https://doi.org/10.1080/23735082.2017.1341053

Kathpalia, S. S., & See, E. K. (2016). Improving argumentation through blog. *System, 58*(2016), 25-36. https://doi.org/10.1016/j.system.2016.03.002

Lam, Y. W., Hew, K. F., & Chiu, K. F. (2018). Improving argumentative writing: Effects of a blended learning approach and gamification. *Language Learning & Technology, 22*(1), 97-118. https://dx.doi.org/10125/44583

Latifi, S., Noroozi, O., Hatami, J., & Biemans, H. J. A. (2021). How does online peer feedback improve argumentative essay writing and learning? *Innovations in Education and Teaching International, 58*(2), 197-208. http://doi.org/10.1080/14703297.2019.1687005

Lin. K. Y., Hong, H.-Y., & Chai, C. S. (2014). Development and validation of the knowledge-building environment scale. *Learning and Individual Differences, 30*, 124-132. https://doi.org/10.1016/j.lindif.2013.10.018

Liu, F., & Stapleton, P. (2014). Counterargumentation and the cultivation of critical thinking in argumentative writing: Investigating washback from a high-stakes test. *System, 45*, 117-128. http://doi.org/10.1016/j.system.2014.05.005

Liu, F., & Stapleton, P. (2020). Counterargumentation at the primary level: An Intervention study investigating the argumentative essay writing of second language learners. *System, 89*(2020), 102-198. https://doi.org/10.1016/j.system.2019.102198

Matos, F. (2021). Collaborative writing as a bridge from peer discourse to individual argumentative writing. *Reading and Writing, 2021*(34), 1321-1342. https://doi.org/10.1007/s11145-020-10117-2.

Means, M. L., & Voss, J. F. (1996). Who reasons well? Two studies of informal reasoning among children of different grade, ability, and knowledge levels. *Cognition and Instruction, 14*, 139-178. https://doi.org/10.1207/s1532690xci1402_1

Mitchell, S. (2000). Putting argument into the mainstream. In S. Mitchell & R. Andrews (Eds.), *Learning to argue in higher education* (pp. 146–154). Boyton/Cook.

Muller Mirza, N., & Perret-Clermont, A. (Eds.) (2009). *Argumentation and education: Theoretical foundations and practices.* Springer.

Nussbaum, E. M., & Kardash, C. M. (2005). The Effects of goal instructions and texts on the generation of counterarguments during writing. *Journal of Educational Psychology, 97*(2), 157-169. http://doi.org/10.1037/0022-0663.97.2.157

Osborne, J. (2010). Arguing to learn in science: The Role of collaborative, critical discourse. *Science, 328*(5977), 463-466. http://doi.org/10.1126/science.1183944

Osborne, J., Simon, S., Christodoulou, A., & Howell-Richardson, K. (2013). Learning to argue: A Study of four schools and their attempt to develop the use of argumentation as a common instructional practice and its impact on students. *Journal of Research in Science Teaching, 50*(3), 315-347. https://doi.org/10.1002/tea.21073

Popper, K. (1972). *Objective knowledge: An Evolutionary approach*. Clarendon Press.

Prata, M. J., de Sousa, B., & Oliveira, A. L. (2019). Cooperative methods and self-regulated strategies development for argumentative writing. *Journal of Educational Research, 112* (1), 12-27. https://doi.org/10.1080/00220671.2018.1427037

Putra, J. W. G, Teufel, S., & Toknaga, T. (2021). Annotating argumentative structure in English-as-a-Foreign-Language learner essay. *Natural Language Engineering*, 1-27. http://doi.org/10.1017/S1351324921000218

Rapanta, C., M. Garcia-Mila, & Gilabert, S. (2013). What is meant by argumentative competence? An Integrative review of methods of analysis and assessment in education. *Review of Education research, 83*(4), 483-520. http://doi.org/10.3102/0034654313487606

Resnick, L. B., Asterhan, C. S. C., & Clarke, S. N. (Eds.). (2015). *Socializing intelligence through academic talk and dialogue*. American Educational Research Association. https://doi.org/10.3102/978-0-93530 2-43-1.

Sawyer, R. K. (2007). *Group genius: The Creative power of collaboration*. Basic Books.

Scardamalia, M. (2004). CSILE/Knowledge forum. In *Education and technology: An encyclopedia* (pp. 183-192). ABC-CLIO.

Scardamalia, M., & Bereiter, C. (1991). Higher levels of agency for children in knowledge building: A Challenge for the design of new knowledge media. *The Journal of the Learning Science, 1*(1), 37-68.

Scardamalia, M., & Bereiter, C. (2003). *Knowledge building*. In J. W. Guthrie (Ed.), *Encyclopedia of Education* (2nd ed., pp. 1370-1373). Macmillan.

Scardamalia, M., & Bereiter, C. (2006). Knowledge building: Theory, pedagogy, and technology. In R. K. Sawyer (Eds.), *Cambridge handbook of the learning sciences* (pp. 97-118). Cambridge University Press.

Simon, S. (2008). Using Toulmin's argument pattern in the evaluation of argumentation in school science. *International Journal of Research and Method in Education, 31*(3), 277-289. http://doi.org/10.1080/17437270802417176

Siqin, T., Van Aalst, J., & Chu, S. K. W. (2015). Fixed group and opportunistic collaboration in a CSCL environment. *International Journal of Computer-Supported Collaborative Learning, 10*, 161-181. https://doi.org/10.1007/s11412-014-9206-7

Sun, Y., Zhang, J., & Scardamalia, M. (2010). Knowledge building and vocabulary growth over two years, Grades 3 and 4. *Instructional Science, 38*, 147-171.

Thompson, I., & Wittek, L. (2016). Writing as mediational tool for learning in the collaborative composition of text. *Learning, Culture and Social Interaction, 11*, 85-96. http://doi.org/10.1016/j.lcsi.2016.05.004

Tomele, G. (2015). The Multi-structural model of speech and language development in the aspect of holistic approach. In *Society. Integration. Education. Proceedings of the International Scientific Conference* (Vol. 3, pp. 222-232). https://doi.org/10.17770/sie2015vol3.392

Toplak, M. E., West, R. F., & Stanovich, K. E. (2013). *Assessing the development of rationality.* In H. Markovits (Ed.), The developmental psychology of reasoning and decision-making (pp. 7-35). Psychology Press.

Toulmin, S. (1958). *The Uses of argument.* Cambridge University Press.

Wingate, U. (2012). Argument! Helping students understand what essay writing is about. *Journal of English for Academic Purposes, 11*(1), 145-154. https://doi.org/10.1016/j.jeap.2011.11.001

Wolfe, C. R. (2011). Argumentation across the curriculum. *Written Communication, 28*(2), 193-219. https://doi.org/10.1177/0741088311399236

Wu, Y.-T., & Wang, L.-J. (2016). Research trends in technology-enhanced knowledge building pedagogies: a review of selected empirical research from 2006 to 2015. *Journal of Computer Education 3*(3), 353-375.

Wu, Y.-T, Wang, L.-J., & Cheng, T.-Y. (2017). A Preliminary study of university students' collaborative learning behavior patterns in the context of online argumentation learning activities: the role of idea-centered collaborative argumentation instruction. In B. K. Smith, M. Borge, E. Mercier, & K. Y. Lim (Eds.), *Making a Difference: Prioritizing Equity and Access in CSCL, 12th International Conference on Computer Supported Collaborative Learning (CSCL)* (Vol. 2, pp. 640-643). International Society of the Learning Sciences. https://dx.doi.org/10.22318/cscl2017.98

Zhang, J., Scardamalia, M., Reeve, R., & Messina, R. (2009). Designs for collective responsibility in knowledge-building communities. *The Journal of the Learning Sciences, 18*(1), 7-44. https://doi.org/10.1080/10508400802581676

Elliott, R. (2023). The Demographics of Student Device Ownership: An Examination of the Personal Computing Ecosystems of Students in Higher Education. *Educational Technology & Society, 26*(3), 129-140. https://doi.org/10.30191/ETS.202307_26(3).0010

The Demographics of Student Device Ownership: An Examination of the Personal Computing Ecosystems of Students in Higher Education

Rob Elliott
Indiana University Purdue University Indianapolis, USA // elliott@iupui.edu

(Submitted May 17, 2022; Revised August 19, 2022; Accepted October 1, 2022)

ABSTRACT: Higher education has become dependent on the use of digital materials, which may include texts, audiovisual content, and software applications. Because students in higher education are largely responsible for providing the computing devices they are required to use to interact with their digital course materials, instructors and instructional designers are often unaware of the personal computing ecosystems in use by their students. This study describes a large-scale survey of student ownership and use of computing devices at a large public university in the midwestern United States. The results demonstrate that students generally have access to devices that allow them to engage with their digital course materials, but age and demographic factors correlated with socioeconomic status appear to impact the type and quality of devices owned. The study also shows that students have access to a variety of device types and that most students perform their computing tasks on a single screen. Understanding the personal computing ecosystems of students will allow instructors and instructional designers to develop course materials that are accessible to students on the devices in use and can inform the decision-making process when an institution considers adoption of new learning technologies. This data can also be used as a foundation for future studies that examine the influence of a student's technology access and ownership on their academic outcomes.

Keywords: Computer-supported collaborative learning, Distance learning/education, Mobile computing, Technology enhanced language learning

1. Introduction

Higher education has become inextricably intertwined with technology (Becker et al., 2017; Gierdowski, 2019b) and students are increasingly bringing more personal technology with them to school (Gierdowski et al., 2020). Cosentino (2020) notes that "…we come to knowledge by engaging with digital technologies that are embedded and embodied, and that extend our cognition" (p. 14). In most cases in higher education in the United States, the students themselves are responsible for supplying the computing devices needed to take advantage of the digital technologies in use in higher education today. The use of computing devices allows students to perform academic work unbounded by classrooms and at the time and place of their choosing (Kukulska-Hulme & Traxler, 2019). Therefore, it is imperative that educators understand the contexts in which students experience their course materials, which are referred to here as the students' "computing ecosystem." If the computing devices through which the digital content is delivered cannot support the activities in which the students must engage, then the benefits of the technology-enhanced learning are diminished or lost (Taylor et al., 2006).

The transition of physical classroom interactions and media into their digital equivalents is both intrinsic, as students see the value of using technology to support their studies (Gierdowski, 2019b), as well as extrinsic, as most institutions have adopted learning management systems (LMS) for communication and organization of instructional materials (Pomerantz et al., 2018). Pomerantz et al. (2018) note that 99% of institutes of higher education (IHEs) have adopted LMS platforms for communication and organization of instructional materials. Further, 88% of faculty were found to use at least some features of the LMS, which indicates they expect students to engage with their courses via some form of computing device. Even prior to the COVID-19 pandemic, most students reported that an LMS was used for "…most or all of their courses" (Galanek et al., 2018, p. 5). IHEs have become replete with technology for even the most mundane classroom tasks (Becker et al., 2017; Gierdowski, 2019b).

Because of this reliance on digital activities in support of academic coursework, a computing device that allows a student to engage with the LMS and other internet and communication technologies (ICTs) used for instruction at an IHE is a de facto, if not outright, requirement (Reisdorf et al., 2020). Indeed, students express a preference for digital learning materials and increased faculty use of technology (Gierdowski, 2019b). Therefore, it is imperative for IHEs and instructors to understand the personal technology owned and used by students – their personal computing ecosystems. Selwyn (2010) argues that "…greater attention now needs to be paid to how digital technologies are actually being used – for better and worse – in 'real-world' educational settings" (p. 66).

To understand this, Allen (2016) notes that faculty and institutions must become familiar with the technology that students use for their academic work. Although the need to understand the technology in use by students has long existed, the transition to online learning due to the COVID-19 pandemic exposed several gaps related to the personal technology in use by students (Gierdowski, 2021; Jaggars et al., 2021).

The so-called "digital divide" is incorrectly thought of as a binary distinction between those who have technology and those who do not (OECD, 2006). More in-depth research into the digital divide generally considers three strata of divisions (Ferreira et al., 2021). The first is access to computing devices and the connectivity required to use them (Deursen & Dijk, 2019; OECD, 2006). The second-level divide examines the computer and information literacy needed to make use of one's computing ecosystem (Dijk, 2006; Hargittai, 2002). Finally, there is a tertiary level that researches an individual's ability to use the computing devices and connectivity to achieve specific outcomes (Cohron, 2015; Rowsell et al., 2017). Obviously, one cannot bridge either the second or third level of the digital divide without the devices and internet access afforded in the first level and IHEs in the United States generally operate on the assumption that their students have sufficient computing access and connectivity to complete their academic requirements (Brooks et al., 2020). However, the quality of the devices in use by individuals can vary wildly (Deursen & Dijk, 2019) and students frequently report difficulties related to connectivity when trying to conduct academic work (Galanek et al., 2018; Gierdowski, 2021; Means & Neisler, 2020). Because of the disparities in student experience related to personal technology, it is important to have a broader understanding of student access to computing devices and connectivity as well as their positive and negative experiences in their everyday use.

This study describes the results of a comprehensive survey of students at a large, multi-campus, public institution of higher education in the midwestern United States which sought to quantify the personal technology those students own and use to complete their collegiate academic work. The purpose of this study is to answer two broad research questions:

- What is the computing ecosystem in use by students at a large public institution of higher education in the United States?
- How do the computing ecosystems of students differ between demographic subgroups?

The results of this study are intended to inform instructors and instructional designers about the personal computing devices in use by their students and provide insight to the potential discrepancies between different subgroups of students with whom they engage. This study should also be useful to the decision-makers at IHEs who are responsible for the evaluation and selection of learning technologies on an institutional scale so that they are able to evaluate the technologies in the context of what students are able to use.

2. Literature review

Students are required to engage with a menagerie of course materials in multiple digital formats and modalities. Traditional textbooks are being replaced or supplemented by electronic texts, interactive digital content, and open educational resources (OER), which is distributed digitally (Moro, 2018; Seaman & Seaman, 2019). Courses have grown beyond the physical confines of the classroom and often include audiovisual content either produced or recommended by the instructor (Brame, 2016). Hybrid and online learning opportunities, which were already on the rise prior to the COVID-19 pandemic, have significantly increased since 2020 and have resulted in additional digital resources for students, including live online interactions and recordings of lectures and classroom sessions (MacKay, 2019). Beyond merely accessing course materials, students are also required to use computers for summative coursework, including proctored exams, digital presentations, and research papers (Schoonenboom, 2012). Students are also occasionally expected to make use of platforms with no physical equivalent, such as social media (Farkas, 2012).

The incorporation of digital artifacts and activities with traditional classroom interaction has been labeled "blended learning" (Owston, 2018), but instructors predominantly use the technology for administrative purposes to improve their efficiency and primarily for the top-down dissemination of academic materials to students (Mpungose & Khoza, 2020; Torrisi-Steele & Drew, 2013). However, because the of transition of physical classroom artifacts into their digital equivalents is so commonplace, Pomerantz et al. (2018) argue that "…it may be time to stop considering trivial uses of online tools (such as using an LMS to post a course syllabus) as worthy of qualifying a course as 'blended' " (p. 4). Laurillard (2005) uses the term "e-learning" to describe the "…shar(ing) of resources across networks," which allows for "…greater flexibility of provision in time and place" (p. 72).

Although it has been demonstrated that students are increasingly bringing more personal technology with them for use in higher education (Gierdowski et al., 2020), detailed information about that technology can be difficult to find. Studies regarding the academic impact of a specific technology will often include details about the particular device(s) of focus but rarely offer a broad picture of student device ownership (Chen et al., 2002; Kenny et al., 2009; Margaryan et al., 2011). EDUCAUSE, a higher education technology advocacy group, conducts a longitudinal study to learn more about the interaction of students in higher education and technology (Galanek et al., 2018; Gierdowski, 2019b; Gierdowski et al., 2020). However, the EDUCAUSE reports, while valuable, do not provide a specific breakdown of the ownership of computing devices. The annual National Survey of Student Engagement (Indiana University Bloomington School of Education, n.d.) included an optional "Learning with Technology" module that was discontinued in 2020. The University of Central Florida (UCF) Center for Distributed Learning conducts a semi-annual survey of student mobile device ownership that reveals near-universal ownership of smart phones but varying levels of tablet ownership and does not include information about the computers used by the students (UCF Center for Distributed Learning, 2018). The Pew Research Center regularly produces reports on computing device ownership of the US population as a whole (Anderson, 2015), but rarely focuses on college students (Smith et al., 2011). Therefore, it is often difficult for researchers to find comprehensive statistics about the computing devices in use by students in higher education. The purpose of this study is to provide a foundational overview of device ownership to assist in the development of additional research questions and allow researchers to delve further into the relationship between a student's personal technology and their success at meeting learning outcomes.

It can be argued that information about student device ownership is critical in the context of understanding student engagement with their digital course materials and their overall academic performance. For example, a 2020 study notes that approximately 20% of students struggle with the technology they have at their disposal and that those with lower-quality computers have lower GPAs (Gonzales et al., 2020). Jaggars et al. (2021) found that 8% of students reported hardware or software issues that were serious enough to disrupt their academic work during the transition to online learning due to COVID-19. These findings support the notion that populations with lower socioeconomic status (SES) have higher barriers to reliable technology and connectivity (Banerjee, 2020; Bell et al., 2022; Gonzales, 2014; Mark et al., 2017) and those technical difficulties are harder to overcome for students of lower SES (Bernhaupt et al., 2020). However, SES disparities are not the only explanation for differences in student performance. Research has shown that students with inadequate computing resources demonstrate worse academic performance than their peers even when controlling for SES factors (Reisdorf et al., 2020). Siani (2017) notes that pedagogy based on the assumption of students' personal technology ownership must consider the digital divide between students who possess and can competently use devices (the first and second level digital divides) and those who cannot.

The access to and quality of a student's computing devices has been shown to have a cascade effect on their academic engagement and performance. Students' ability to cope with technical difficulties (Pituch & Lee, 2006) and access to technical support (Sánchez et al., 2013) can improve their computing self-efficacy. Improved self-efficacy, in turn, contributes to a students' perceived usefulness of a learning technology (Huang, 2020). Alsabawy et al. (2016) found that the quality of the IT infrastructure of an e-learning system – which, we must assume, includes the devices on which the student is engaging – also has a direct effect on the student's perceived usefulness of that system. Perceived usefulness, defined by Davis (1989) as "…the degree to which a person believes that using a particular system would enhance his or her job performance" (p. 320) has frequently been used to gauge students' willingness to use personal technology for academic work in higher education (Alsabawy et al., 2016; Lai et al., 2012). Measuring usefulness related to information technology has been expanded into the Technology Acceptance Model (TAM) (Davis et al., 1989), which has similarly been used to examine the factors that determine a students' adoption of technology to engage in their academic work. A meta-analysis of this work was performed by Granić and Marangunić (2019).

Consequently, students with limited or inadequate access to computing devices can demonstrate lower levels of computer and digital literacy than their peers (Hargittai, 2002; Hargittai, 2010). A students' lack of digital literacy can negatively affect their usage of e-learning technologies, such as an LMS (Oz et al., 2015). The task-technology-fit (TTF) model posits that there is a connection between a user's experience with a technology and its subsequent utilization (Goodhue & Thompson, 1995). McGill and Klobas (2009) applied the TTF model to LMS usage and found that students' perception of a fit between the task and the technology used to accomplish the task had a significant effect on their attitude toward LMS use. If a student's technology cannot accomplish the tasks they are expected to perform, their attitude toward the entire academic endeavor may suffer.

Although laptops and computers are most often the focus of research related to student computing devices, studies show that students are increasingly using handheld mobile devices – such as phones and tablets – to engage with their academic work (Cross et al., 2019; Gierdowski et al., 2020; Magda et al., 2020; Seilhamer et al., 2018). Therefore, understanding the mobile devices students own and use may be important when making a comprehensive examination of student academic engagement through technology. Other work investigates the impact of allowing students to view multiple digital inputs simultaneously and suggests that students retain information better when they have more viewable screen area or multiple screens (Hsu et al., 2012; Lanir et al., 2010; Lanir et al., 2013; Miller et al., 2020). Given the variety of computing devices and peripherals used in the conduct of academic work, it stands to reason that a broad overview of the hardware in use by students is an important factor in any examination of a student's relationship with digitalized course materials. This study provides such an overview as both a model for similar studies as well as a foundation for further research in the use of personal technology in higher education.

3. Methods

The study was conducted by anonymous survey at a major public university system in the US Midwest with over 90,000 students. The university has nine physical campuses (including a core residential campus, a large urban semi-residential campus, five regional campuses, and two satellite campuses) and a slate of exclusively online programs. All students in an online program are also assigned to one of the nine physical campuses and thus appear in the results for that campus type. Eligible participants included anyone over the age of 18 who was enrolled at any campus during the Fall 2021 semester. Enrolled students of any level (undergraduate, graduate, and professional) were asked to participate.

Purposive sampling of the students was performed to ensure that a variety of disciplines and campuses are represented. Approximately 30% of the students at every campus were included in the initial sample. Invitations were made via emails to university-assigned email addresses and delivered through the Qualtrics survey management system. Survey questions were drawn from multiple sources, including previous ad hoc interactions with students, and inspired by widely cited studies and reports (Cross et al., 2019; Galanek et al., 2018; Gierdowski et al., 2020; Gikas & Grant, 2013). In addition to asking about the computing devices owned, students were also asked to report the number of external monitors used with their computers, thus allowing the separate calculation of the number of device screens and the total number of computing screens. The survey also asked students about demographic factors that are not part of the institutional demographic profile, such as number of hours worked per week, living situation, and availability of high-speed internet.

The survey responses were then paired with institutional demographic records about each respondent (including their major, age, and enrollment status). The survey results were analyzed with both descriptive and inferential statistics (Leedy & Ormrod, 2016) using Python 3.8 and the statsmodels package. Crosstabulations and multiple regression analyses (Flick, 2015) were performed to reveal relationships between demographic and ownership factors (i.e., do students from regional campuses own desktop computers at a higher rate than students at the larger campuses?)

3.1. Definitions in this study

For the purposes of this study, computing devices are divided into three distinct categories. The term "mobile device" is defined as one that uses a mobile operating system (such as iOS, Android, or iPadOS). This includes smart phones and tablets such as iPads. "Computers," therefore, are defined as devices that run a full version of an operating system (including laptops and hybrid tablet/laptops such as the Microsoft Surface). Chromebooks represent a hybrid device in that they use ChromeOS – a limited, semi-mobile operating system – but have the physical affordances of a traditional laptop computer (Pegoraro, 2021). Thus, "Chromebooks" are treated as a third class of device in this study.

4. Results

All students over the age of 18 who were enrolled in the Fall 2021 semester at any campus or online program of the university was eligible to participate in the anonymous student survey. A sample of 26,966 eligible students was created. From this sample, 149 individuals were removed because of invalid email addresses. This sample represented approximately 30% of all eligible participants at the university. The survey began on October 11, 2021 and remained open for 28 days until November 8, 2021. Survey invitations were delivered via Qualtrics' internal email distribution system. All participants received an initial invitation to participate followed by two reminder messages sent at weekly intervals. The 2,146 responses that were received resulted in a response rate of

8.0%. Of these responses, 2,041 included information about the computing devices the students owned or regularly used.

4.1. Demographic overview of responses

The demographics of the respondents were diverse and roughly aligned with the population of the university. 55.9% of the respondents were 18-21 years old, 16.2% were 22-25, and 14.9% were 26-35. White students were the largest ethnic group (68.0%), followed by Hispanic/Latino (8.0%), Black/African American (5.9%), and Asian (5.3%). Females were somewhat overrepresented and made up 65.0% of respondents. International students provided 7.9% of the responses. Students at all points in their plans of study were represented: 22.5% were undergraduates in their first year, 15.6% in their second, 14.0% in their third, and 20.2% were in their fourth. Graduate and professional students account for the remaining 27.7% of responses. Over 140 different plans of study were included in the sample; the five most common majors by the respondents include Finance, Psychology, Biology, Computer Science, and Nursing. 46.5% of the responses were from the core campus, 29.2% from the urban campus, 22.0% from the regional campuses, and 2.3% were from the satellite campuses. 22.1% of respondents reported being a first-generation student, which is an indicator of parental education level. 21.5% of respondents were eligible for US federal Pell Grants, which is an indicator of financial need. Full-time students were 80.4% of responses, and 78.1% of students lived off-campus. Students with a private, unshared study space made up 78.6% of responses. Over 91% of students reported having high-speed internet at their place of residence. Finally, students had a wide range of work commitments: 31.6% of students do not work, 12.6% of students work 1-10 hours per week, 22.4% work 11-20 hours, 16.2% work 21-39 hours, and 17.2% work 40 hours or more.

4.2 Summary of device ownership

When inquiring about ownership, the decision was made to use more inclusive language; thus, students were asked to identify each of the computing devices that they “own or regularly use.” The results appear in Table 1.

Table 1. Computing device ownership (*n* = 2041)

Device	# of students who own	Percentage of respondents
Smart phone	2,006	98.3%
Laptop	1,887	92.5%
Tablet	718	35.2%
Desktop computer	513	25.1%
E-reader	125	6.1%
Chromebook	115	5.6%
“Basic” mobile phone	10	0.5%

E-readers (such as a Kindle) and basic mobile phones provide limited access to the learning technologies in place at the study site. Therefore, the remainder of the analysis will focus on ownership and use of five computing devices: smart phones, tablets, laptop and desktop computers, and Chromebooks.

Table 2 demonstrates that laptops are the most common computer in the study and are owned (or used) by 92.4% of respondents. Although 25.1% of respondents report owning or using a desktop computer, just 2.3% of respondents exclusively use a desktop. Laptops are also the only non-mobile computing device for 68.3% of respondents. Computer and Chromebook ownership is high but not universal – 2.6% of respondents report that they have neither a computer nor a Chromebook.

Table 2. Computer and Chromebook ownership (*n* = 2041)

Device	% of students who own	% of students w/o another computer
Laptop computer	92.4%	68.3%
Desktop computer	25.1%	2.3%
Chromebook	5.6%	2.2%

A plurality of respondents – 48.8% – own just two of the five computing devices included in the study. The most common combination of computing devices, as depicted in Table 3, is a smart phone and laptop, which is the computing ecosystem used by 44.5% of respondents in the study. While few students (2.2%) appear to be dependent on a Chromebook in lieu of a laptop or desktop computer a gender disparity exists. 2.3% of respondents who identify as female use a smart phone and Chromebook combination, compared to just 0.8% of respondents who identify as male.

Table 3. Common device ownership combinations (*n* = 2041)

Device	# of students	% of sample
Smart phone, laptop	908	44.5%
Smart phone, laptop, tablet	465	22.8%
Smart phone, desktop, laptop	245	12.0%
Smart phone, desktop, laptop, tablet	188	9.2%
Smart phone, Chromebook	36	1.8%
Smart phone	31	1.5%

Table 4 shows that the majority of respondents report use a single screen when using a computer or Chromebook, regardless of the type of device in use. For laptops and Chromebooks that single screen is the screen attached to the device itself. Although most respondents who use desktop computers do use a single screen, respondents with desktops are more likely to have two screens than those using a laptop or a Chromebook. These numbers may also be influenced by the wide range of possible desktop monitor sizes, which were not included in the survey; a single external monitor could effectively have the same physical dimensions of two or more smaller screens.

Table 4. Number of computer and Chromebook screens (*n* = 2041)

Device	# (%) of students who own	% of owners using		
		1 screen	2 screens	3+ screens
Laptop computer	1887 (92.4%)	76.3%	20.1%	3.4%
Laptop computer only	1393 (68.3%)	79.6%	18.1%	2.3%
Desktop computer	513 (25.1%)	56.5%	38.4%	5.1%
Chromebook	115 (5.6%)	79.0%	18.4%	2.6%
Chromebook only	44 (2.2%)	86.4%	13.6%	0.0%

4.3. Influence of demographics on device ownership

Logistic regression analysis was used to discover predictors of demographic categories that may have an impact on the ownership and computing devices and working environments in use by respondents. The odds ratio was calculated for each explanatory demographic variable for three of the most owned or used computing devices (desktop computer, laptop computer, and tablet). Further, odds ratios were also calculated for three computing ecosystems where a student uses a single screen for their academic work. The result of this analysis is found in Table 5.

A student's age was found to be a significant predictor in both desktop computer and tablet ownership, as well as the use of a laptop with no monitor or a desktop computer with a single monitor. Because age was such a statistically significant predictor for these variables, logistic regression analysis was conducted a second time while controlling for the age of the respondent. The results of the secondary analysis generally confirmed the first but revealed additional statistically significant interactions between demographics and device ownership and use. When controlling for age, international students were found to be less likely to own a desktop computer (OR 0.66, 95% *CI*: 0.44-0.98, $p < .05$). Also when controlling for age, students at regional campuses were found to be less likely to own a laptop computer (OR 0.60, 95% *CI*: 0.40-0.89, $p = .01$). First-generation students, regardless of age, were also found to be less likely to own a laptop computer (OR 0.61, 95% *CI*: 0.43-0.87, $p < .01$), as were Pell-eligible students (OR 0.67, 95% *CI*: 0.46-0.97, $p < .05$).

Laptop ownership varied based on several demographic categories. Black/African American respondents were 41% as likely to own a laptop when compared to their white peers, and just 28% of Hispanic/Latino respondents owned a laptop in comparison to white respondents. Accordingly, these two groups showed a significantly increased dependence on a Chromebook in lieu of a computer: Black/African American students were over four times as likely and Hispanic/Latino students were seven times as likely to own a Chromebook only and not a laptop or a desktop computer. A respondent's first-generation status also appeared to influence their laptop ownership; first-generation respondents were much less likely to own a laptop than respondents with a family history of college attendance.

Table 4. Demographic factors of device ownership (odds ratios) (*n* = 2041)

Demographic	Desktop	Laptop	Tablet	Chromebook only (no monitor)	Laptop only (no monitor)	Desktop only (one monitor)
Gender (ref. M)						
F	0.34***	1.10	1.36**	2.12	1.75***	0.44*
Age (ref. 18-21)						
22-25	1.45**	1.47	1.27	0.25	0.75*	0.98
26-35	1.84***	1.34	2.09***	0.27	0.41***	0.80
36-45	2.49***	0.77	2.24***	1.31	0.40***	2.04
46-55	2.86***	0.65	2.42***	1.65	0.20***	4.35*
56+	7.32***	0.44	2.23*	1.34	0.11**	7.00*
Ethnicity (ref. White)						
Asian	0.97	0.52	0.91	1.52	0.93	2.62
Black/African American	0.77	0.41*	0.81	4.25**	1.10	2.35
Hispanic/Latino	0.76	0.28***	0.93	6.99***	0.77	2.90
Two or more	0.69	0.72	0.93	1.02	1.30	1.16
Unknown	1.62	0.50	1.06	-	0.19*	10.78**
Enrollment Type (ref. FT)						
Part-Time (<12 cr. Hours)	1.83***	0.70	1.55***	1.43	0.45***	3.03**
Campus Type (ref. Core)						
Urban	1.06*	1.03	1.01	0.98	0.85	1.06
Regional	1.35*	0.59	1.09	2.65*	0.88	1.95
Satellite	1.34	0.57**	0.97	3.20	0.87	5.84
International student						
Yes	0.75	0.78	1.19	-	0.75	0.80
First-generation student						
Yes	1.15	0.60**	0.93	1.79	1.12	1.03
Pell-eligible						
Yes	0.88	0.71	0.74**	2.33*	1.30*	0.88
Classification (ref. UG)						
Graduate	1.56***	1.43	1.75***	0.33*	0.48***	1.02
Professional	0.69	3.59	3.09***	-	0.45***	-
Living status						
Off-campus	1.72***	1.01	1.22	1.05	0.69	1.63
Hours Worked/wk (ref. 0)						
1-10	1.06	1.09	1.27	1.26	1.07	0.55
11-20	1.14	1.23	0.89	0.70	1.02	0.94
21-39	1.28	1.05	1.21	1.38	0.93	0.65
40+	2.21***	1.01	1.97***	1.11	0.38***	1.44
Private study space						
No	0.73*	1.12	0.84	0.99	1.61***	0.64

Note. $^{*}p < .05$; $^{**}p < .01$; $^{***}p < .001$.

The use of a laptop without an external monitor varied based on several demographic conditions. As the age of the respondent increases, so too does the likelihood that the respondent uses at least one external monitor with their laptop computer. Similarly, graduate and professional students (who tend to be older) were far less likely than undergraduate students to use a laptop without an external monitor. Respondents without a private study space were 61% more likely to use a laptop without an external monitor, which indicates that they may have more transient habits related to where they perform their academic work.

5. Discussion and implications

Overall ownership and access to computing devices for students appears to be quite high. Most students (98.3%) own a smart phone and 97.4% report owning or using either a computer or a Chromebook. When examining computers and Chromebooks, laptops were the most widely adopted device type (92.4%), which aligns with previous surveys of a similar nature (Galanek et al., 2018; Gierdowski, 2019a; Reisdorf et al., 2020). However, 2.2% of students use a Chromebook as their primary computer and 2.6% of students do not have a primary computer or Chromebook. Laptop ownership is slightly – not significantly – correlated with age, but students who are Black/African American, Hispanic/Latino, or first-generation students show statistically significantly lower likelihood of laptop ownership. These demographics are often correlated with lower SES and, consequently, lower rates of technology ownership (Banerjee, 2020; Gonzales, 2014; Reisdorf et al., 2020).

Tablets appear to be a supplementary computing device; they are owned by just 35.1% of students and fewer than 1.0% of students report using a tablet in lieu of a computer or Chromebook. Female students tend to adopt tablets at higher rate than males, and ownership of tablets increases with a students' age and number of hours worked. Students who are eligible for Pell Grants are less likely to own a tablet, but the devices are more prevalent with graduate and professional students. These findings are supported in previous studies that show cost is a major factor in computer equipment decisions for students (Reisdorf et al., 2020) and that students prioritize the purchase of a phone and a computer over that of a tablet (Elliott, 2022).

The use of Chromebooks by students in higher education is a relatively recent phenomenon but one that should be of particular interest to IHEs. The increased presence of Chromebooks on campus can be likely be attributed to several factors: cost, access, and prior use. Chromebooks are generally less expensive than other laptop computers which may indicate why they are more popular with Pell-eligible students, students of ethnicities correlated with lower SES, and the regional campuses. However, the COVID-19 pandemic and the transition to online learning resulted in the increased use of computer-based education for students in K-12 schools. The younger students in this survey – those that were most recently in K-12 schools – may simply be more comfortable with the Chromebook interface based on their previous experience or may even be using the same devices that carried them through the end of their secondary school experience during the COVID-related transition to online learning.

Although Chromebooks allow access to any compatible web-based learning technology, their limited operating system can present barriers when students are required to install or use specialized software for their academic work. MATLAB, for example, cannot easily be installed directly on a Chromebook (Vivirito, 2013; Mitchell, 2018). Students without access to a laptop or desktop computer would need to use that application's online portal or the IHE may need to provide access via a virtualized environment. Similarly, students working on a Chromebook would be relegated to using the online version of Microsoft Office products as they cannot install the native applications. This could impact students who are expected to use advanced features that are not available in the web-based Office applications. Situations such as this will require IHEs to invest in licenses that afford online access, and instructors and instructional designers may have to provide additional or alternative instruction for the use of online or alternative interfaces in addition to that of the standard installed software.

The majority of students report working on a single computer screen when performing their academic work. Of students who own a laptop as their only non-mobile computing device, 79.6% use the laptop screen alone. Students using Chromebooks are even more likely to use the screen of the device without another monitor. Given that students are required to engage with multiple digital materials while conducting their studies – the LMS, electronic texts, and video resources, just to name a few – they may find difficulties in their ability to reference multiple resources simultaneously with such limited viewing area. A student interviewed by EDUCAUSE during the pandemic reported that they had to drop a class because the online course "…required Photoshop, Zoom, and a photo editor app running simultaneously," which the student's computing ecosystem could not support (Gierdowski, 2021). Further study is needed to determine if students' academic performance can be positively impacted with additional monitors or screen area. However, given the high percentage of students who have access to a computer or Chromebook, institutions may wish to consider replacing some of their existing computing infrastructure to provide students with peripherals (such as docking stations and external monitors) that allow them to enhance their experience when using their personal devices. Regardless, institutions, instructors, and designers should be aware of the limited viewing area available to most students when designing their instructional resources and curriculum. Some students may not be able to fully engage in activities without moving between windows or enlarging the digital content.

6. Conclusion

This study analyzes the results of a survey of students at a large, public IHE in the midwestern United States. The purpose of the study is to provide an up-to-date overview of the personal computing ecosystems in use by students in higher education so that instructors, instructional designers, and institutions can align their pedagogies with the technology available to their students. The results demonstrate that students generally own or have access to an array of computing devices with which they can conduct academic work, but a student's specific computing ecosystem is correlated with several demographic factors. Smart phones and laptop computers are the students' primary computing devices and most students (79.6%) who use a laptop do so without an external monitor. Students in demographic groups that correspond with lower SES show lower rates of laptop ownership and a higher likelihood of using a Chromebook to complete their academic work. First-generation and Pell-eligible students have fewer computing resources than their counterparts, as do students who study at the smaller satellite campuses in the system in this study. Age was found to be a significant factor in the computing devices owned by students; as students get older they appear to acquire additional or improved computing resources, including tablets, desktop computers, and external monitors.

6.1. Future work

Although this research provides a thorough examination of the personal computing ecosystems at the study site, it makes no attempt to compare the students' personal devices on their academic achievement. Student devices may not necessarily be the determinant between academic outcomes, but the demographic breakdowns detailed here should provide a foundation for work related to the academic achievement of specific subgroups of students. Additionally, while information about ownership and access to devices is important, more work is needed to study the impact of device ownership in conjunction with the environments in which those devices are used. It would be worthwhile to compare students with transient computing ecosystems and habits to their peers with more robust (but less portable) setups to determine if this is a factor in academic outcomes.

Very little work exists that compares learning outcomes between students using different numbers of screens. The few studies that directly compare single- and multiple-monitor configurations show that the use of multiple monitors may reduce cognitive load (Miller et al., 2020) and support improved student learning (Hsu et al., 2012; Lanir et al., 2010). This is an area ripe for study.

Finally, the delineation between mobile devices, "computers," and Chromebooks in this study should further the notion that the student use of non-traditional devices in higher education is increasing to the point where their use must be seriously considered by IHEs and instructional designers. Studies that compare student outcomes when using these different device families – whether or not the study controls for demographic factors – may be able to reveal any potential disadvantages students face when using different types of primary computing devices. Chromebooks in particular may pose a challenge to students. Students in introductory courses may be using web-based versions of applications whose features and behavior can differ from the instructional materials provided. Chromebooks may also hinder students in advanced courses that require significant computing resources and/or specialized software.

6.2. Limitations of the study

Many of the demographic categories described in this study, including gender, race/ethnicity, and international student status, came from institutional data that was collated with the survey results. This institutional data, unfortunately, has a limited number of categories available – particularly for the gender and race/ethnicity categories. Survey respondents were not provided the opportunity to self-identify their gender identity, race, or ethnicity which prevented a more granular analysis of individual student responses for these demographic items. The author sincerely regrets that this is the case.

References

Allen, B. J. (2016). Optimizing technology's promise. *EDUCAUSE Review*, *51*(6). https://er.educause.edu/articles/2016/10/optimizing-technologys-promise

Alsabawy, A. Y., Cater-Steel, A., & Soar, J. (2016). Determinants of perceived usefulness of e-learning systems. *Computers in Human Behavior, 64*, 843–858. https://doi.org/10.1016/j.chb.2016.07.065

Anderson, M. (2015). *The Demographics of device ownership*. Pew Research Center. https://www.pewresearch.org/internet/2015/10/29/the-demographics-of-device-ownership/

Banerjee, M. (2020). An Exploratory study of online equity: Differential levels of technological access and technological efficacy among underserved and underrepresented student populations in higher education. *Interdisciplinary Journal of E-Skills and Lifelong Learning*, *16*, 93–121. https://doi.org/10.28945/4664

Becker, S. A., Brown, M., Cummins, M., & Diaz, V. (2017). *NMC horizon report: 2017 higher education edition*. The New Media Consortium. https://www.learntechlib.org/p/174879/

Bell, T., Aubele, J. W., & Perruso, C. (2022). Digital divide issues affecting undergraduates at a Hispanic-serving institution during the pandemic: A Mixed-methods approach. *Education Sciences*, *12*(2), 115. https://doi.org/10.3390/educsci12020115

Bernhaupt, R., Mueller, F. "Floyd," Verweij, D., Andres, J., McGrenere, J., Cockburn, A., Avellino, I., Goguey, A., Bjørn, P., Zhao, S. (Shen), Samson, B. P., Kocielnik, R., Petro, G., Gonzales, A., & Calarco, J. (2020). "Out of luck": Socio-economic differences in student coping responses to technology problems. *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (pp. 1–10). https://doi.org/10.1145/3313831.3376156

Brame, C. J. (2016). Effective educational videos: Principles and guidelines for maximizing student learning from video content. *CBE Life Sciences Education*, *15*(4), es6. https://doi.org/10.1187/cbe.16-03-0125

Brooks, D. C., Grajek, S., & Lang, L. (2020, April 9). *Institutional readiness to adopt fully remote learning*. Educause Review. https://er.educause.edu/blogs/2020/4/institutional-readiness-to-adopt-fully-remote-learning

Chen, Y.-S., Kao, T.-C., Sheu, J.-P., & Chiang, C.-Y. (2002). A Mobile scaffolding-aid-based bird-watching learning system. *Proceedings of the IEEE International Workshop on Wireless and Mobile Technologies in Education* (pp. 15–22). https://doi.org/10.1109/wmte.2002.1039216

Cohron, M. (2015). The Continuing digital divide in the United States. *The Serials Librarian*, *69*(1), 77–86. https://doi.org/10.1080/0361526x.2015.1036195

Cosentino, E. (2020). *Artifacts and affordances* (Vol. 198). Routledge; eBook Collection (EBSCOhost). https://doi.org/10.1007/s11229-019-02297-4

Cross, S., Sharples, M., Healing, G., & Ellis, J. (2019). Distance learners' use of handheld technologies: Mobile learning activity, changing study habits, and the "place" of anywhere learning. *The International Review of Research in Open and Distributed Learning*, *20*(2). https://doi.org/10.19173/irrodl.v20i2.4040

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, *13*(3), 319. JSTOR. https://doi.org/10.2307/249008

Davis, F. D., Bagozzi, R. P., & Warshaw, P. R. (1989). User acceptance of computer technology: A Comparison of two theoretical models. *Management Science*, *35*(8), 982–1003. https://doi.org/10.1287/mnsc.35.8.982

Deursen, A. J. van, & Dijk, J. A. van. (2019). The First-level digital divide shifts from inequalities in physical access to inequalities in material access. *New Media & Society*, *21*(2), 354–375. https://doi.org/10.1177/1461444818797082

Dijk, J. A. G. M. van. (2006). Digital divide research, achievements and shortcomings. *Poetics*, *34*(4–5), 221–235. https://doi.org/10.1016/j.poetic.2006.05.004

Elliott, R. (2022). *The Use of mobile devices for formal learning in higher education: Investigating student behaviors and expectations* (Unpublished doctoral dissertation). Indiana University, Bloomington, IN. https://hdl.handle.net/2022/27468

Farkas, M. (2012). Participatory technologies, pedagogy 2.0 and information literacy. *Library Hi Tech*, *30*(1), 82–94. ABI/INFORM Collection; ProQuest One Academic. https://doi.org/10.1108/07378831211213229

Ferreira, D., Vale, M., Carmo, R. M., Encalada-Abarca, L., & Marcolin, C. (2021). The Three levels of the urban digital divide: Bridging issues of coverage, usage and its outcomes in VGI platforms. *Geoforum*, *124*, 195–206. https://doi.org/10.1016/j.geoforum.2021.05.002

Flick, U. (2015). *Introducing research methodology: A Beginner's guide to doing a research project*. SAGE Publications Ltd.

Galanek, J. D., Gierdowski, D. C., & Brooks, D. C. (2018). *ECAR study of undergraduate students and information technology (2018)*. EDUCAUSE Center for Analysis and Research. https://tacc.org/sites/default/files/documents/2018-11/studentitstudy2018_0.pdf

Gierdowski, D. C. (2019a). *ECAR study of community college students and information technology, 2019*. ECAR. https://www.educause.edu/ecar/research-publications/ecar-study-of-community-college-students-and-information-technology/2019/introduction-and-key-findings

Gierdowski, D. C. (2019b). *ECAR study of undergraduate students and information technology, 2019*. ECAR. https://library.educause.edu/resources/2019/10/2019-study-of-undergraduate-students-and-information-technology

Gierdowski, D. C. (2021). *Student experiences with connectivity and technology in the pandemic*. EDUCAUSE. https://www.educause.edu/ecar/research-publications/2021/student-experiences-with-connectivity-and-technology-in-the-pandemic/introduction-and-key-findings

Gierdowski, D. C., Brooks, D. C., & Galanek, J. (2020). *2020 Student technology report: Supporting the whole student*. EDUCAUSE Center for Analysis and Research. https://library.educause.edu/resources/2020/10/2020-student-technology-report-supporting-the-whole-student

Gikas, J., & Grant, M. M. (2013). Mobile computing devices in higher education: Student perspectives on learning with cellphones, smartphones & social media. *Internet & Higher Education*, *19*, 18–26. https://doi.org/10.1016/j.iheduc.2013.06.002

Gonzales, A. L. (2014). Health benefits and barriers to cell phone use in low-income urban U.S. neighborhoods: Indications of technology maintenance. *Mobile Media & Communication*, *2*(3), 233–248. https://doi.org/10.1177/2050157914530297

Gonzales, A. L., Calarco, J. M., & Lynch, T. (2020). Technology problems and student achievement gaps: A Validation and extension of the technology maintenance construct. *Communication Research*, *47*(5), 750–770. https://doi.org/10.1177/0093650218796366

Goodhue, D. L., & Thompson, R. L. (1995). Task-technology fit and individual performance. *MIS Quarterly*, *19*(2), 213. JSTOR. https://doi.org/10.2307/249689

Granić, A., & Marangunić, N. (2019). Technology acceptance model in educational context: A Systematic literature review. *British Journal of Educational Technology*, *50*(5), 2572–2593. https://doi.org/10.1111/bjet.12864

Hargittai, E. (2002). Second-level digital divide: Differences in people's online skills. *First Monday*, *7*(4). https://doi.org/10.5210/fm.v7i4.942

Hargittai, E. (2010). Digital na(t)ives? Variation in internet skills and uses among members of the "net generation"*. *Sociological Inquiry*, *80*(1), 92–113. https://doi.org/10.1111/j.1475-682x.2009.00317.x

Hsu, J.-M., Chang, T.-W., & Yu, P.-T. (2012). Learning effectiveness and cognitive loads in instructional materials of programming language on single and dual screens. *Turkish Online Journal of Educational Technology*, *11*(2), 156–166.

Huang, C.-H. (2020). The Influence of self-efficacy, perceived usefulness, perceived ease of use, and cognitive load on students' learning motivation, learning attitude, and learning satisfaction in blended learning methods. *2020 3rd International Conference on Education Technology Management* (pp. 29–35). https://doi.org/10.1145/3446590.3446595

Indiana University Bloomington School of Education. (n.d.). *Evidence-based improvement in higher education*. https://nsse.indiana.edu/

Jaggars, S. S., Motz, B. A., Rivera, M. D., Heckler, A., Quick, J. D., Hance, E. A., & Karwisch, C. (2021). *The Digital divide among college students: Lessons learned from the COVID-19 emergency transition*. Midwestern Higher Education Compact. http://files.eric.ed.gov/fulltext/ED611284.pdf

Kenny, R. F., Neste-Kenny, J. M. C. V., Park, C. L., Burton, P. A., & Meiers, J. (2009). Mobile learning in nursing practice education: Applying Koole's FRAME model. *Journal of Distance Education*, *23*(3), 75–96.

Kukulska-Hulme, A., & Traxler, J. (2019). Design principles for learning with mobile devices. In H. Beetham & R. Sharpe (Eds.), *Rethinking pedagogy for a digital age: Designing for 21st century learning* (pp. 181-196). Routledge. https://doi.org/10.4324/9781351252805

Lai, C., Wang, Q., & Lei, J. (2012). What factors predict undergraduate students' use of technology for learning? A Case from Hong Kong. *Computers & Education*, *59*(2), 569–579. https://doi.org/10.1016/j.compedu.2012.03.006

Lanir, J., Booth, K. S., & Hawkey, K. (2010). The Benefits of more electronic screen space on students' retention of material in classroom lectures. *Computers & Education*, *55*(2), 892–903. https://doi.org/10.1016/j.compedu.2010.03.020

Lanir, J., Booth, K. S., & Wolfman, S. A. (2013). An Observational study of dual display usage in university classroom lectures. *Human–Computer Interaction*, *28*(4), 335–377. https://doi.org/10.1080/07370024.2012.697037

Laurillard, D. (2005). E-learning in higher education. In P. Ashwin (Ed.), *Changing higher education: The Development of learning and teaching* (pp. 71–84). Routledge.

Leedy, P. D., & Ormrod, J. E. (2016). *Practical research: Planning and design*. Pearson.

MacKay, J. R. D. (2019). Show and "tool": How lecture recording transforms staff and student perspectives on lectures in higher education. *Computers & Education*, *140*, 103593. https://doi.org/10.1016/j.compedu.2019.05.019

Magda, A. J., Capranos, D., & Aslanian, C. B. (2020). *Online college students 2020: Comprehensive data on demands and preferences* (p. 58). Wiley Education Services. https://edservices.wiley.com/ocs2020/

Margaryan, A., Littlejohn, A., & Vojt, G. (2011). Are digital natives a myth or reality? University students' use of digital technologies. *Computers & Education*, *56*(2), 429–440. https://doi.org/10.1016/j.compedu.2010.09.004

Mark, G., Fussell, S., Lampe, C., Hourcade, J. P., Appert, C., Wigdor, D., & Gonzales, A. (2017). Technology maintenance. *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, 289–294. https://doi.org/10.1145/3025453.3025475

McGill, T. J., & Klobas, J. E. (2009). A Task–technology fit view of learning management system impact. *Computers & Education*, *52*(2), 496–508. https://doi.org/10.1016/j.compedu.2008.10.002

Means, B., & Neisler, J. (2020). *Suddenly online: A National survey of undergraduates during the COVID-19 pandemic*. Digital Promise. https://digitalpromise.org/wp-content/uploads/2020/07/ELE_CoBrand_DP_FINAL_3.pdf

Miller, R. A., Stenmark, C. K., & Ittersum, K. van. (2020). Dual computer displays reduce extraneous cognitive load. *Journal of Computer Assisted Learning*, *36*(6), 890–897. https://doi.org/10.1111/jcal.12442

Mitchell, R. (2018, January 20). *Matlab on an Acer Chromebook?* MathWorks. https://www.mathworks.com/matlabcentral/answers/378123-matlab-on-an-acer-chromebook

Moro, J. (2018). The Emergence of digital course materials in higher education and their effectiveness in teaching and engaging students. *Publishing Research Quarterly*, *34*(3), 417–429. https://doi.org/10.1007/s12109-018-9594-z

Mpungose, C. B., & Khoza, S. B. (2020). Postgraduate students' experiences on the use of Moodle and Canvas learning management system. *Technology, Knowledge and Learning*, 1–16. https://doi.org/10.1007/s10758-020-09475-1

OECD. (2006, January 4). *Digital divide definition*. https://stats.oecd.org/glossary/detail.asp?ID=4719

Owston, R. (2018). Empowering learners through blended learning. *International Journal on E-Learning*, *17*(1), 65-83.

Oz, H., Demirezen, M., & Pourfeiz, J. (2015). Digital device ownership, computer literacy, and attitudes toward foreign and computer-assisted language learning. *Procedia - Social and Behavioral Sciences*, *186*, 359–366. https://doi.org/10.1016/j.sbspro.2015.04.028

Pegoraro, R. (2021, March 25). *What is a Chromebook?* U.S. News & World Report. https://www.usnews.com/360-reviews/technology/laptops/what-is-a-chromebook

Pituch, K. A., & Lee, Y. (2006). The Influence of system characteristics on e-learning use. *Computers & Education*, *47*(2), 222–244. https://doi.org/10.1016/j.compedu.2004.10.007

Pomerantz, J., Brown, M., & Brooks, D. C. (2018). *Foundations for a next generation digital learning environment: Faculty, students, and the LMS*. ECAR. https://library.educause.edu/-/media/files/library/2018/1/ers1801.pdf

Reisdorf, B. C., Triwibowo, W., & Yankelevich, A. (2020). Laptop or bust: How lack of technology affects student achievement. *American Behavioral Scientist*, *64*(7), 927–949. https://doi.org/10.1177/0002764220919145

Rowsell, J., Morrell, E., & Alvermann, D. E. (2017). Confronting the digital divide: Debunking brave new world discourses. *The Reading Teacher*, *71*(2), 157–165. https://doi.org/10.1002/trtr.1603

Sánchez, R. A., Hueros, A. D., & Ordaz, M. G. (2013). E-learning and the University of Huelva: A Study of WebCT and the technological acceptance model. *Campus-Wide Information Systems*, *30*(2), 135–160. https://doi.org/10.1108/10650741311306318

Schoonenboom, J. (2012). The Use of technology as one of the possible means of performing instructor tasks: Putting technology acceptance in context. *Computers & Education*, *59*(4), 1309–1316. https://doi.org/10.1016/j.compedu.2012.06.009

Seaman, J. E., & Seaman, J. (2019). *Inflection point: Educational resources in U.S. higher education, 2019*. https://www.bayviewanalytics.com/reports/2019inflectionpoint.pdf

Seilhamer, R., Chen, B., Bauer, S., Salter, A., & Bennett, L. (2018). Changing mobile learning practices: A Multiyear study 2012–2016. *Educause Review*. https://er.educause.edu/articles/2018/4/changing-mobile-learning-practices-a-multiyear-study-2012-2016

Selwyn, N. (2010). Looking beyond learning: Notes towards the critical study of educational technology. *Journal of Computer Assisted Learning*, *26*(1), 65–73. https://doi.org/10.1111/j.1365-2729.2009.00338.x

Siani, A. (2017). BYOD strategies in higher education: Current knowledge, students' perspectives, and challenges. *New Directions in the Teaching of Physical Sciences*, *12*(1). ERIC. https://eric.ed.gov/?id=EJ1231963

Smith, A., Rainie, L., & Zickuhr, K. (2011, July 19). *College students and technology*. https://www.pewresearch.org/internet/2011/07/19/college-students-and-technology/

Taylor, J., Sharples, M., O'Malley, C., Vavoula, G., & Waycott, J. (2006). Towards a task model for mobile learning: A Dialectical approach. *International Journal of Learning Technology*, *2*(2/3), 138. https://doi.org/10.1504/ijlt.2006.010616

Torrisi-Steele, G., & Drew, S. (2013). The Literature landscape of blended learning in higher education: The Need for better understanding of academic blended practice. *International Journal for Academic Development*, *18*(4), 1–13. https://doi.org/10.1080/1360144x.2013.786720

UCF Center for Distributed Learning. (2018, January 1). *2018 mobile survey report*. https://digitallearning.ucf.edu/msi/research/mobile/survey2018/

Vivirito, H. (2013, August 5). *How could I run MatLab on a Chromebook?* MathWorks. https://www.mathworks.com/matlabcentral/answers/84055-how-could-i-run-matlab-on-a-chromebook

Chen, L. (2023). Students' Learning Styles and Preferences in a Gamification-enhanced Partially Flipped Classroom: A Q-Methodology Study. *Educational Technology & Society*, *26*(3), 141-154. https://doi.org/10.30191/ETS.202307_26(3).0011

Students' Learning Styles and Preferences in a Gamification-enhanced Partially Flipped Classroom: A Q-Methodology Study

Liwen Chen
UWE Undergraduate Double Degree Program in Business Administration, Chung-Hua University, Taiwan, R.O.C. // lwchen@g.chu.edu.tw

(Submitted May 23, 2022; Revised September 3, 2022; Accepted October 1, 2022)

ABSTRACT: The aim of this Q-study was to identify and categorize learners' learning styles and preferences with regard to the incorporation of gamification-enhanced activities in a partially flipped gamified classroom during a Taiwan university eighteen week's Introduction to Marketing course. Q-methodology was used because it identifies assorted viewpoints subjectively and analyzes them statistically. Twenty-six students were surveyed and asked to rank thirty statements according to their perception of the teaching method used. A factor analysis and a correlation test were used to identify both the factors involved and the individuals with whom they were highly correlated. Three factors were identified: Factor A – Engaged Achiever, Factor B – Self-motived Explorer, and Factor C – Interactive Designer, each of which represented participants with similar perceptions. These multiple learning styles and perspectives present both challenges and opportunities in business education.

Keywords: Partially flipped classroom, Gamification, Badges, Q-methodology

1. Introduction

There is an increasing awareness in various educational institutions of the implementation of Flipped Classroom (FC) methods. FC refers to the practice of designing course materials, such as instructional videos, text-based materials, and online exercises outside class, while devoting time in class to a wide variety of interactive learning activities (DeLozier & Rhodes, 2016; Lo & Hew, 2020). Previous researchers have reported that the use of FC methods can promote students' academic performance, enhance their interaction, improve their attendance, and cultivate positive attitudes (Chen et al., 2014). Bhagat et al. (2016) point out that FC methods can also help low-achievers to improve because their teachers tend to pay more attention to them.

However, in contrast, Cabi (2018) observed that FC did not have a significant effect on academic enhancement. Some students also expressed concerns about having to do homework before classes (Chen et al., 2015). Liou et al. (2016) found that, as videos in FC were not interactive, other technologies needed to be incorporated in order to enhance learning. According to Sun and Wu (2016), while FCs result in better learning outcomes, there seems to be no significant "between-group" difference in teacher-student interactions. Consequently, other activities, such as integrated classroom polling systems, mobile game-based learning, or multimedia Learning Management Systems (LMSs) were needed to boost interactivity.

Despite the recent promotion of FCs by many educators and practitioners, Ye et al. (2018) found that appropriate teaching strategies were required, either before or during classroom activities, to help learners to organize their own studies and to augment higher-order cognition. The use of gamification in higher education as a means of engaging, retaining, and motivating students is advocated in much of the literature (Hew et al., 2016). However, there is little published data on the effect of gamification on FC activities. Nevertheless, as there is evidence that some learners skip out-of-class activities or pre-class video lectures, when gamification is not used, perhaps they would be more effectively engaged by its incorporation.

Recent FC developments have led to an interest in its partial use; hence, it is not necessary to apply it to entire courses. Selectively flipping the most difficult information makes the workload involved in developing flipped material less overwhelming for instructors because there is no need to flip the entire course, which gives them additional choices to modify the delivery of the course information based on its level of complexity (Strelan et al., 2020; Urquiza-Fuentes, 2020). To date, little is known of learners' subjective experiences and preferences in partially flipped classrooms in Taiwan and, since they have not been studied, the effectiveness of gamification remains to be seen. Consequently, because FC teaching is increasingly being used in higher education, it is essential for outcomes of the initial application of partial FCs to be investigated, particularly in terms of students' responses to out-of-class learning, in order to determine how to best incorporate them into the learning process.

How students learn, and especially their learning styles, based particularly on understanding individual learning preferences, has received considerable critical attention from researchers in numerous fields (Hassan et al., 2019). However, learning-style instruments cannot provide an in-depth understanding of learning skills as many of them are quantitatively-based. Consequently, Q-Methodology, which incorporates quantitative and qualitative methods, is used in this study to categorize students' opinion of a gamified, partially flipped classroom at a Taiwanese university. The study is structured as follows: (i) theoretical and empirical background, (ii) review of the relevant literature, (iii) description of the methodological approach, (iv) results, and (v) a conclusion, which includes a discussion of the results, a description of the limitations of the study, and some recommendations for further investigations in this field.

2. Literature review

2.1. Use of gamified flipped classrooms in higher education

A gamified flipped higher education classroom operates in combination with game-based learning, which appears to be a promising method of instruction; however, gamification and game-based learning are not interchangeable, either in their definition or application. Game-based learning refers to the inclusion of games in learning activities to achieve instructional goals, whereas gamification refers to the inclusion of game elements or mechanics, such as "points." (numerical evidence of performance), "avatars" (visual representations of players' characters), "virtual goods" (online assets with perceived gaming value), "leaderboards" (direct comparison of expertise), or "badges" (visual representations of accomplishments) to enhance students' interest and motivation through competition in a non-game context (Buckley & Doyle, 2017; de-Marcos et al., 2016; Subhash & Cudney, 2018; Sun & Hsieh, 2018).

However, specific game elements, such as badges, have been found to undermine students' motivation instead of improving it (Facey-Shaw et al., 2020; Muilenburg & Berge, 2016). Furthermore, Buckley and Doyle (2017) stated that active, or global, learning style-orientated individuals had a positive perception of gamification, as did extraverts, whereas conscientious types were less motivated by it.

In summary, the primary aim of using gamified learning in an FC classroom is to integrate selected game-based elements with a view to increasing students' motivation, higher-level thinking, and self-efficacy.

2.2. Self-determination theory

The Self-determination Theory (SDT) is based on individuals' personality and motivation to indicate how they interact with and rely on society. Motivation can be either intrinsic or extrinsic. Intrinsic motivation is based on autonomy, competence, relatedness, self-determination, well-being, and engagement; it can refer to participating in educational activities that are enjoyable, interesting, appealing, and exciting. In contrast, extrinsic motivation drives individuals to constantly strive for rewards, avoid criticism or punishment, and have a diminished desire for autonomy. Hence, it lies on a continuum of identifiable behavioral regulations, such as external, introjected and integrated, which reflect the degree to which behavior creates a sense of self (Deci & Ryan, 1985, 2002).

The SDT contains two forms of motivation, namely, controlled and autonomous. The former involves behaving for external reasons, such as gaining rewards and avoiding punishment or guilty feelings, which creates a sense of obligation and stress. This kind of behavior is likely to be maintained for as long as rewards are on offer, but it will probably discontinue without external reinforcements. On the other hand, autonomous motivation drives self-determined behavior for intrinsic ends, such as through choice and interest; therefore, this kind of behavior will probably persist, even if a reward seems to be unlikely (Deci & Ryan, 2002; Hagger et al., 2014).

The SDT has recently been implemented in several gamification and FC studies in order to examine learning motivation, engagement, and performance (Kuo & Chuang, 2016; Tinati et al., 2017; Thai et al., 2017). It is important to note that the level of motivation affects the extent to which students will engage in, and persist with, certain behavior.

2.3. Learning styles

Several inventories have been made over the past thirty years, which suggest that learning styles have matured and been permeated with persuasive variables related to learning strategies that reflect their developers' diverse backgrounds. This has resulted in various theories, instruments, and empirical works based on different theories. For instance, Curry (1987) used an onion metaphor to categorize three layers of learning styles, (i) cognitive personality elements, (ii) information-processing, and (iii) instructional preferences, while Riding and Cheema (1991) used over thirty theoretical divisions they called "cognitive, or learning, styles." Neil Fleming's VARK model, which categorizes instructional preferences, classifies four kinds of learning preferences based on sensory pathways: (i) visual, (ii) aural, (iii) read/write, and (iv) kinesthetic (Aldosari et al., 2018). Students are classified into four dimensions in the Felder-Silverman learning-style model (FSLSM) based on the bipolar categories of sensory/intuitive, visual/verbal, active/reflective, and sequential/global: (1) Perception: how information is perceived (sensory-intuitive), (2) Input: how information is presented (visual-verbal), (3) Processing: how information is processed (active-reflective), and (4) Understanding: how information is understood (sequential-global) (Felder & Silverman, 1988). The Index of Learning Styles (ILS), developed by Felder and Soloman (1997) based on the FSLSM, is influenced by Jung's psychological types, the Myers-Briggs Type Indicator, and Kolb's learning processes. Hu et al. (2021) suggest that the FSLSM is a better model for a technology-enhanced learning environment since it includes essential cognitive learning theories and practices.

Wang et al. (2004) describe the student learning process as more complex in a web-based education environment than in a traditional classroom; as a result, a traditional learning-style typology may be inappropriate for a web-learning environment. They established a novel typology of web-based learning styles that included aggressive knowledge-seekers, active participants, silent cultivators, and heavy sleepers. In summary, although numerous models characterize preferred learning styles, it is imperative for instructors to recognize individual and group differences.

3. Methodology

3.1. Procedures

This study was conducted at a university in Hsinchu, Taiwan, during an eighteen-week Introduction to Marketing course. Prior to the research, a semester-long pilot test was also undertaken in the Human Resource Management course. The Introduction to Marketing course, which is an introduction to the basic principles and application of marketing practices, is a required component of a Department of Technology Management program. The course involved three one-hour classroom sessions each week. Twenty-six undergraduates whose major is Technology Management participated in the study. Their ages ranged from 19 to 20 years old, and they were evaluated based on sixteen written assignments and case studies, seven sessions of gamification-enhanced activities in a flipped classroom, and mid-term and final exams.

This course was based on a "partial-flip" approach, in which a flipped classroom format only accounted for a portion of the class time (7 times). Text-based lecture notes and pre-recorded multimedia micro-lectures were delivered via the school's Moodle e-learning system. Formal teams, each comprised of between two and four trainees, were assigned problem-solving and decision-making tasks, as well as being required to complete a peer review. A Moodle course management platform was designed to support the uploading of course resources and activities that included quizzes, assignments, and a digital badge display. The flipped classroom, based on Deci and Ryan's (1985, 2002) Self-Determination Theory, incorporated active learning strategies, while gamification-enhanced activities were designed to fulfill students' psychological needs of autonomy, relatedness, and competence (see Figure 1).

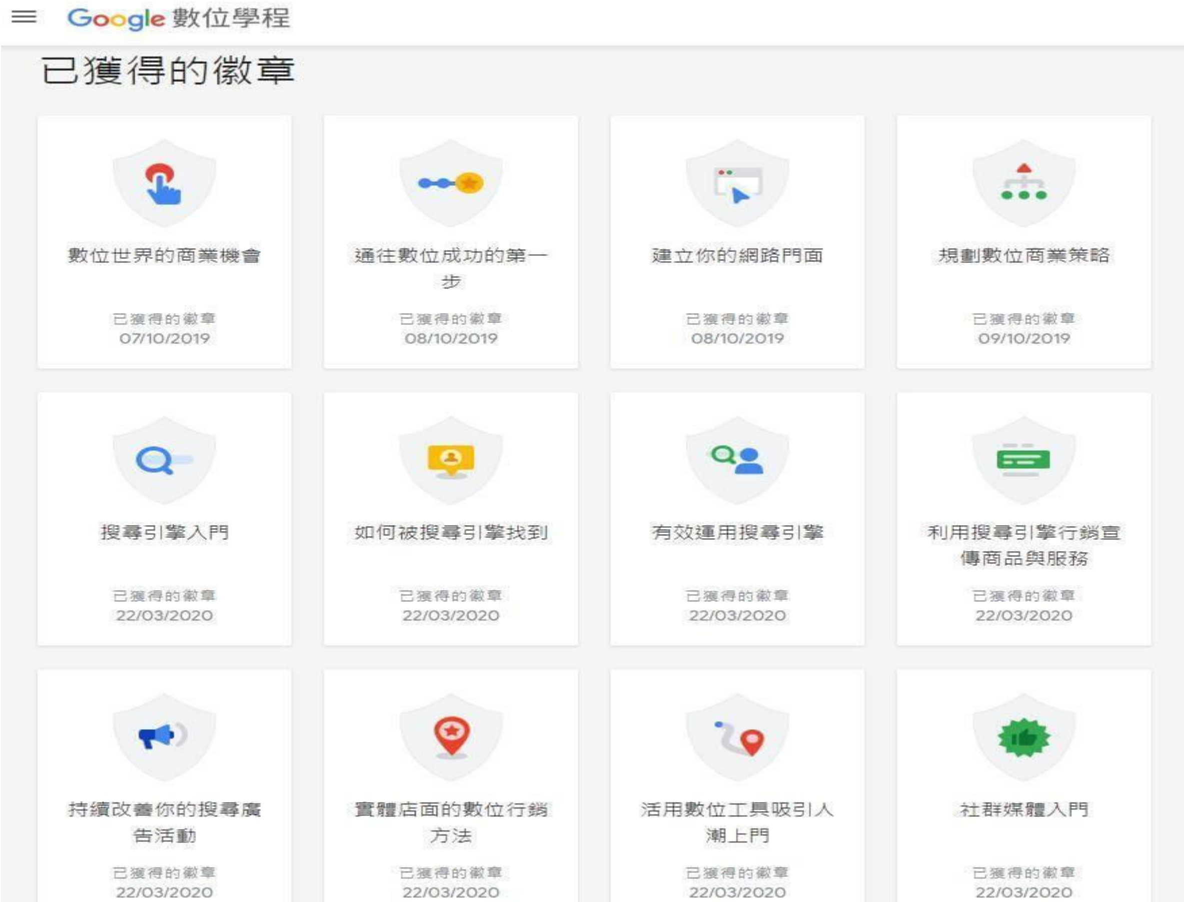
It is important to note that Hew et al. (2016) and Denden et al. (2021) have validated the efficiency of the proposed SDT based gamification design. Four types of badges were used, (i) autonomy-based, such as "early-bird," (ii) relatedness-based, such as "reply warrior," (iii) competence-based, and (iv) text-based, where the instructor wrote personal comments on "magic stone" badges based on individuals' performance.

There were two levels of competence-based badges: (i) apprentice, and (ii) knowledge expert (see Figure 1). Students who logged into Moodle e-Learning to access the course materials before class received an autonomy-based badge. Those who participated in discussion forums received a relatedness-based badge, and those who completed flipped activities, such as self-directed quizzes, received a competence-based badge.

Figure 1. Badges earned in class

Students who completed all twenty-six modules earned twenty-six badges (see Figure 2), and were also awarded Google's Fundamentals of Digital Marketing certificate via self-directed learning.

Figure 2. Badges from Google Certification



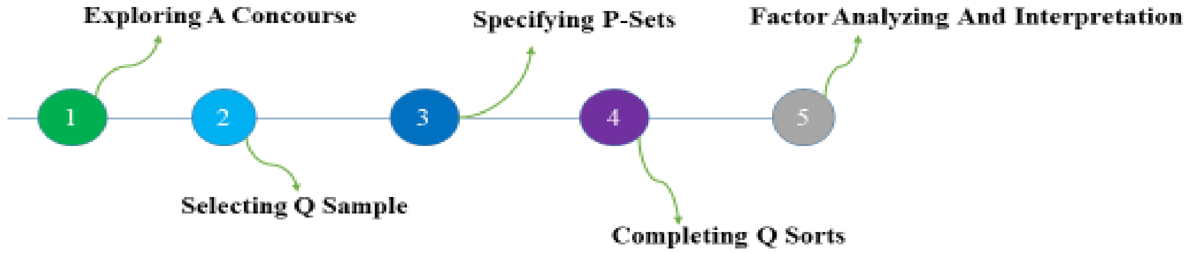
3.2. Measuring subjectivity

Q-methodology was chosen as the research tool to distinguish the students' learning styles and acquire a deeper understanding of their perception and how they learn. Q-methodology is a distinctive set of psychometric and operational principles which, when combined with statistical applications of correlational and factor-analytic techniques, provide researchers with a systematic and rigorously quantitative procedure for investigating the subjective components of human behavior; however, it does not identify causes or generalize demographic variables in a large population (Brown, 1993). A small sample was used for this study, and since Q-methodology effectively explains the main participants' perspectives, their number was deemed to be unimportant. Research subjects in Q-methodology are often chosen due to theoretical or pragmatic considerations (i.e., by convenience sampling or particular relevance to the topic). According to Stephenson (1935), Q methodology was designed to give a small number of individuals a large number of tests, test items, or responses, in contrast to the majority of quantitative research, which gives a large number of participants a small number of tests. As a result, the value of Q-methodology lies in discovering the opinions and understanding of groups of participants; therefore, while most studies are effective with forty to sixty participants, some require far fewer. (Valaitis et al., 2007; Watts & Stenner, 2012). According to Webler et al. (2009), numerous Q-studies have between 12 and 20 participants, who are usually chosen intentionally, purposefully or strategically. Although a small sample size is not an issue in Q-methodology, the participants must be familiar with the topic and have a distinctive opinion of it (Chng et al., 2022; Watts & Stenner, 2012).

3.3. Data collection

Q-methodology has five stages: (i) exploring a list of items defining the topic's perspectives (the concourse), (ii) selecting the Q-sample (Q-statements or Q-sets) by refining those items to provide a well-rounded research perspective of the topic, (iii) specifying a P-set (participants), which is non-random and comprised of relevant individuals, (iv) completing the Q-sort, where each participant is identified as a sorter who orders Q-statements and administers the Q-sort, and (v) a factor analysis and interpretation (see Figure 3). The correlation between the sorts, the factor analysis, and the factor score, is calculated during stage (v). The factors related to the groups of participants can be interpreted when the Q-factor analysis is complete (Brown 1993), McKeown and Thomas (2013), Chen et al. (2015) for a further theoretical explanation.

Figure 3. Stages of Q-methodology



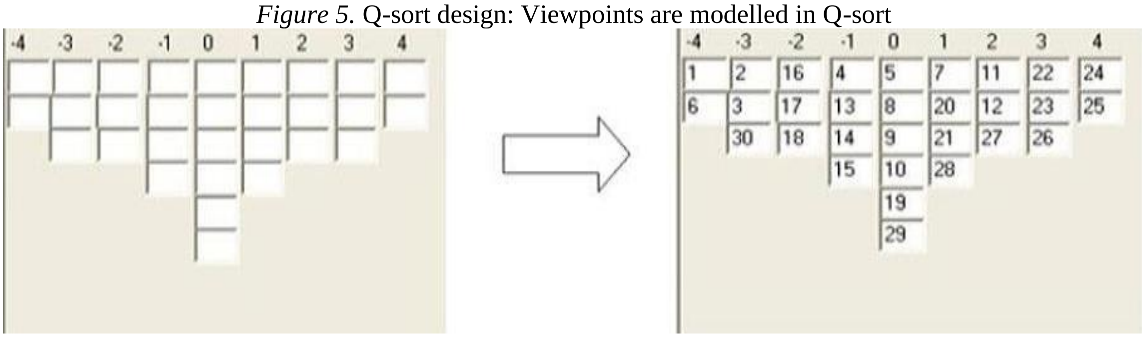
After completing the course, the twenty-six students took part in a Q-study, beginning with semi-structured interviews to respond to the research question, "Tell me about your learning experience. How did you learn that?" A representative sample of thirty statements from the interviews containing key ideas was used to develop the research instrument (i.e., Q-set) (see Figure 4). The process of creating the Q-set consists of gathering distinctive assertions, thoughts, or concepts related to the subject, preferably up to saturation point. The sampling may involve a literature review, preliminary data collection (e.g., interviews), or searching for other publicly-available resources. Photographs and other images may also be used. Opinion statements can be gathered from any primary or secondary source where the issue of interest is discussed. The collected statements are then reduced and refined (e.g., by grouping similar ideas together) to produce a manageable Q-set.

Fisher's variance design is the most formal way to ensure the comprehensiveness of the Q-sample, with equal numbers of statements chosen from each cell of a theoretically informed two-dimensional matrix. Some Q methodologists advocate a more liberated, creative approach that is focused on understanding and representing the statement population as a whole (Brown, 1993). Different from the present study, Hall et al. (2013) adapted the existing instrument of Soloman-Felder ILS as the Q-statement of a Q-method study in an introductory geographic information systems course. Fisher's variance design was not used to structure the Q-sample in the present study because it was not designed to select statements to meet a predetermined quota. Instead, the final Q-set of statements was chosen through a content analysis that characterized aspects of technology, content, and the teacher and participants. Each participant was given a questionnaire in which they were asked to share their thoughts about this novel pedagogy. They were asked to rank thirty statements into nine categories, ranging from Most Disagreeable (-4) to Most Agreeable (+4) (see Figure 5), providing their opinions in Q-sorts on the answer sheet, without bias and with equal treatment for disagreements and agreements (Watts & Stenner, 2012).

Figure 4. Examples of statement

Item	Statement
1	I am passionate about progressing towards Google certification.
2	I gained valuable expertise via this approach.
3	I am very happy to know that I obtained the rare “Magic Stone Badge” since it is very scarce and hard to get.
4	I appreciate this way of teaching because I can collect many badges.
5	It is important for me to get more than 20 badges in the Moodle system.

Figure 5. Q-sort design: Viewpoints are modelled in Q-sort



3.4. Data analysis

Data and Q-sorts were entered into the PQMethod (version 2.11) statistical software program, which resulted in different piles of statement numbers. The statements were examined, and various methods of factor rotation and statistical procedures were used to preserve the factor reliability. Correlation, a centroid factor analysis, and judgmental (hand) rotation were used to identify the significant factors in this partial FC context (Watts & Stenner, 2012).

Researchers have shown that the correlation values for test-retest reliability are usually .80 or higher when the same people are given the same instrument (Q-sample) on two separate occasions. Q-methodology has produced similar findings when the same set of statements is used with different person samples, and different Q-samples drawn from the same concourse yield similar results. Since the respondents’ Q-sorts are neither right nor wrong, but constructed by their rank-ordering of self-referent items, validity in line with quantitative research tenets is not a concern in Q-studies (Valenta & Wigger, 1997).

4. Results

To address the research aim, the findings revealed (i) a 3-factor structure – Factors A, B, and C being three types of learning styles (see Table 1), and (ii) because factor scores were used to represent the characteristics of each cluster, the clusters were defined by the uniqueness of statements that were combined to define each factor’s distinctiveness (Table 1), with the first factor describing an engaged achiever (EA – Factor A), the second describing a self-motivated explorer (SME – Factor B), and the third describing an interactive designer (ID – Factor C). These three factors were named by comparing and contrasting the three sets of distinguishing statements, which helped to define and explain the uniqueness of each factor (see Table 2). Those arranged in the most important (+4 and +3) and least important -4 and -3) columns were distinguishing items, while each factor was labeled with a name so that it could be seen in the distinguishing items rankings. The complete pattern of the statements helped to identify the discriminated clusters of learners. The analysis of the Q-sorts revealed three distinctive factors. Twenty-three (88.5%) of the participants’ Q-sorts were divided into these factors and three were found to be statistically insignificant. Twelve (52.17%) participants’ Q-sorts were Factor A, six (26.09%) were Factor B, and five (21.74%) were Factor C (see Table 1).

Table 1. Factor structure of student group (*)

ID	Factor A	Factor B	Factor C
1	.83		
2	.84		
6	.50		
7	.60		
8	.56		
9	.52		
11	.77		
13	.60		
14	.74		
15	.63		
16	.79		
26	.74		
3		.84	
5		.65	
12		.51	
18		.70	
21		.80	
24		.45	
10			.70
17			.42
19			.65
20			.44
22			.48

Note. *Only significant loadings shown ($p < .01$).

4.1. Factor A: Engaged achiever

The Q-sorts of twelve participants: 6 males and 6 females, were significantly loaded, as evidenced by strong positive and negative statements (Table 2). familiarization with the performance evaluation and the badging mechanism pre-course was important (Statement 6), as were positive feelings about achieving badges and Google certification through self-directed learning. Teacher-student interactions were also enhanced (Statements 4, 5, 23, 25) and there was strong disagreement with Statement 14, "I personally prefer to work alone to complete the tasks although this course requires group discussion and collaboration." The mean scores of the mid-term and final exams for the EAs were 91.09 (SD = 8.006) and 86.82 (SD = 6.794), respectively.

4.2. Factor B: Self-motivated explorer

The Q-Sorts of six participants: 5 males and 1 female, were significantly loaded, as evidenced by distinguishing items rankings (Table 2). SMEs, like EAs, indicated the importance of achieving badges and progressing toward Google certification independently (Statements 5, 25). Interestingly, neither EAs nor SMEs expressed a wish to design and issue badges to peers. Both groups indicated that not seeing any students tended to dominate the class discussions. They disagreed that team members usually compete with each other (Statements 9, 11, 21). The SMEs strongly stated that they read pre- course materials and answered fundamental and advanced pre-class questions (Statements 26, 27, 30). The SMEs' mid-term and final exam mean scores were 86 (SD = 9.933) and 81.29 (SD = 6.291), respectively.

4.3. Factor C: Interactive designer

The Q-sorts of five students: 3 males and 2 females, were significantly loaded, as evidenced by their strong positive and negative statements (Table 2). They strongly agreed that they had gained considerable knowledge and were delighted to learn at their own pace (Statements 2, 22). They also believed that the teaching method increases teacher-student interaction (Statement 23). Interestingly, this group wished they could have designed customized badges for their classmates (Statement 21). They described some students as dominating the panel discussions, allowing fewer opportunities for others to participate. They did not feel that team members competed with each other (Statements 9, 10, 11) (Table 2). Unlike the SMEs they described themselves as being unprepared for class by not completing the pre-class questions. They were the only participants who did not have a positive experience with the leaderboard because the accumulated points/badges were displayed publicly (Statements 26, 7). The IDs' mean scores of the mid-term and final exams were 69.2 (*SD* = 10.895) and 71 (*SD* = 7.211), respectively.

Table 2. Statement scores by factor/opinion types

Factor statements		A	B	C
	Factor A: Engaged achiever			
6	I get to know the performance evaluation mechanism and points/badges percentages in each activity before classes.	4		
25	I appreciate gaining extra Google certification via self-directed learning.	4		
4	I appreciate this way of teaching because I can collect many badges.	3		
5	It is important for me to get more than 20 badges in the Moodle system.	3		
23	This approach enhances my interaction with my instructor.	3		
9	Team members usually compete with each other when they join the group discussion.	-4		
10	I sometimes disagree with other group members in a group discussion.	-4		
11	Some members have a dominant voice during panel discussions so that others have fewer opportunities to speak.	-3		
14	I prefer to work alone to complete tasks although this course requires group discussion and collaboration.	-3		
21	I wish I could design my own personalized badge in class and distribute it to my classmates.	-3		
	Factor B: Self-Motivated explorer			
5	It is important for me to get more than 20 badges in the Moodle system.		4	
30	I usually download and read course materials on the Moodle system before class.		4	
25	I appreciate gaining extra Google certification via self-directed learning.		3	
26	I usually answer not only basic questions, but also advanced questions of the pre-class quiz listed on the MOODLE system.		3	
27	I usually answer the basic questions on the Moodle system before classes.		3	
9	Team members usually compete with each other in group discussions.		-4	
10	I sometimes disagree with other group members in a group discussion.		-4	
11	Some members have dominant voices during panel discussions so others have fewer opportunities to speak.		-3	
19	I like the appearance (form, color, style) of the badges		-3	
21	I wish I could design my own personalized badge in class and distribute it to my classmates.		-3	
	Factor C: Interactive designer			
2	I gained valuable expertise via this approach.			4
23	This teaching method increases my interaction with my instructors.			4
11	Some members have dominant voices during panel discussions so others have fewer opportunities to speak.			3
21	I wish I could design my own personalized badge in class and distribute it to my classmates.			3
22	This course is taught via pre-class previews and classroom discussions. I am pleased to learn at my own pace.			3
7	I had a positive experience with the leaderboard since everyone can see the accumulated points/badges publicly. It is important for me to use the leaderboard to interpret my rank relative to that of others.			-4
9	Team members usually compete with each other when they join group discussions.			-4
3	I am very happy to know that I obtained the rare "Magic Stone Badge" since it is very scarce and hard to get.			-3
10	I sometimes disagree with other group members in group discussions.			-3
26	I usually answer not only the basic questions, but also the advanced questions of the pre-class quiz listed on the MOODLE system.			-3

Note. Item rankings: -4 = most unimportant in this sample; 0 = ambivalent; +4 = most important in this sample

4.4. Consensus statements

While those with all three learning styles had opposing views on many issues, there was agreement on a few (Table 3). Consensus statements – those not distinguishing any factor pairs in the three groups – appear in Statements 2, 9, 10, 22, 23, 25.

Table 3. Consensus statements between Factor A, B, and C

Statements	Factors	A	B	C
2	I gained valuable expertise via this approach.	2	1	4
9	Team members usually compete with each other when they join group discussions.	-4	-4	-4
10	I sometimes disagree with other members in group discussions.	-4	-4	-3
22	This course is taught via pre-class previews and classroom discussions. I am glad I can learn at my own pace.	1	1	3
23	This teaching method increases my interaction with my instructor.	3	1	4
25	I appreciate gaining extra Google certification via self-directed learning.	4	3	1

4.5. Learning outcomes

A one-way analysis of variance (ANOVA) was used to determine if there were any significant differences between the students' mid-term and final exam scores and learning outcomes (final exam scores – mid-term scores) in terms of different factors. The results, which are shown in Table 4, indicate a significant difference between the mid-term and final exam scores of students in terms of different factors, but not a significant level of different learning effectiveness. F-values of 9.734 and 9.501 for mid-term and final exam scores respectively, with *p*-values of .001 and .001 for each factor, reached a significant difference level.

Table 4. Multiple comparisons of learning results for Factors A (EAs), B(SMEs), and C(IDs)

Learning results	Group	Number of students	Mean	*SD*	*F-value*	Post hoc comparison
Results of Mid-term Exam	EAs	12	91.09	8.006	9.734^{***}	1 > 3, 2 > 3
	SMEs	6	86.00	9.933		
	IDs	5	69.20	10.895		
Results of Final Exam	EAs	12	86.82	6.794	9.501^{***}	1 > 3
	SMEs	6	81.29	6.291		
	IDs	5	71.00	7.211		
Learning Effectiveness (Midterm vs. Final exam)	EAs	12	-4.27	3.580	2.914	
	SMEs	6	-4.71	5.090		
	IDs	5	1.80	7.791		

Note. $^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$.

A follow-up Scheffe's test was also conducted and the results indicated that the mid-term exam scores of both EAs and SMEs were significantly higher than those of IDs. The final exam scores of EAs were significantly higher than those of IDs. In terms of learning effectiveness, the means of the final exam scores of EAs and SMEs were less than those of their mid-term scores. The means of the final exam scores of IDs were higher than their mid-term scores. The F-value of learning effectiveness was 2.914, and the p-value was 0.077, which did not reach the significant difference level.

5. Discussion and conclusion

5.1. Discussion

5.1.1. Construct whole or partial FC experiences

Previous researchers found that not all students agreed that the FC improved their learning due to cultural values. For instance, Asian students, including Taiwanese, are generally passive in articulating their personal opinions in class; hence, it is challenging to motivate them to engage in a fully-flipped class. The findings of this study show that using two different teaching methods simultaneously, namely, a partially flipped classroom and traditional lecturing, was able to meet the needs of a group of diverse Taiwanese university students in a business education setting. This finding corresponds to that of Waldrop and Bowdon (2016), who also found that partially-flipped teaching appeared to be better than fully-flipped for an entire semester. It is challenging for some adult students to adapt to inverted learning; hence, it may be easier for those who are uncomfortable with technology, or new to flipped learning, to receive only small segments of their course using this method until they become more familiar with it.

5.1.2. Care for learning styles and gender differences

Some conclusions could be drawn to address the research questions from the examination of the three sets of opinions expressed in Factors A, B, and C. To begin with, the EA, SME, and ID groups of participants in this flipped classroom utilized three distinctive learning styles. Although they expressed diverse preferences and opinions of digital badges, they had some themes in common. Interestingly, the EA group contained six of the nine female participants (66.67%), which implies that the perception of gamification may be gender-specific. This finding is consistent with the finding of Koivisto and Hamari (2014) that gamification provides women with greater social benefits, which suggests that instructors and curriculum designers should strive to understand gender differences in respect of the diverse incentives prevalent in flipped classrooms. Such socially-relevant features may be vital, especially for educators who wish to recruit users to help in the design and implementation of gamification. This result differs from the finding of Wang et al. (2004) that there was no significant difference between gender and learning styles in a web-learning environment.

According to Felder and Silverman (1988), active learners prefer to collaborate in teams in order to discuss, question, argue, brainstorm, experiment, or reflect. It was found in the present study that EAs with higher grades interacted more frequently during the course. This finding supports Huang et al.'s (2012) conclusion that the sensory/intuitive dimension of a learning style indirectly predicts the learning performance through the mediation of online participation. This increased online participation results in a better e-learning performance. This finding is consistent with that of Wang et al. (2004). The finding in the present study is aligned with the conclusion of Cela et al. (2015) that learning styles may yield insights, which educators can use to provide opportunities across learning styles and develop opportunities for students to use their individual strengths to improve their learning outcomes.

5.1.3. Challenges to formal and informal learning

Every participant earned all 26 badges and received a personal Google Fundamentals of Digital Marketing certificate. They practiced for hours after class at their own pace and with their own targets without receiving instruction; hence, they took responsibility for their own learning, as well as nurturing an interest in the subject. As Song and Bonk found in 2016, with such a wide range of online resources and emerging technologies, the potential for an increase in informal, self-directed learning is growing. However, the badge-issuing system linked to those goals is regarded as a constraint of freedom in the context of formal and informal learning.

The participants in the Self Motivated Explorer group, who tended to be self-challenging risk-takers, answered the Moodle questions voluntarily in advance of classes; hence, gamification appears to have contributed to better engagement with their studies, which Hamari et al. (2016) also found to be the case. Chen and Chen (2018) observed that some educators divide their course materials into different levels. They assist students to work to the level of their capability by providing all students with the required pre-class materials, but giving higher performers optional learning resources. Most learners in Factor B, which was 83.33% male, appeared to derive more benefit than females from information sources and more demanding work. This revelation may prove to be a critical factor in gamification design and application.

5.1.4. Competition on the leaderboard

All the EAs, SMEs, and IDs strongly believed that digital badges could motivate students to learn. This finding is not consistent with that of Facey-Shaw et al. (2020), who found that badges did not enhance students' intrinsic motivation in an introductory programming course. Leaderboards emphasize the social feature of badges by displaying the ranking of players in descending order. The relationship between gamification, points awarded and the function of leaderboards was highlighted as an extrinsic motivator in past studies due to seemingly enhanced performances as learners saw themselves climbing up the leaderboard (Mekler et al., 2017; Seaborn & Fels, 2015). It was found in this study that members of the Interactive Designer group were the only ones to oppose the idea of displaying their badges publicly; hence, one of the drawbacks of a leaderboard is that it could demotivate some students to the extent that they may leave the game altogether.

However, there are alternatives for providing learners with a better sense of their relative ranking. For instance, instructors could assign each student an online pseudonym at the beginning of the semester, although this strategy should be treated with caution because some researchers have found that anonymity may lead to more negative electronic contributions, causing social "loafing"; i.e., when people are part of a group, they tend to exert less effort. Social loafing can be reduced by providing comparative feedback about each group member's performance, such as displaying the average points/badges earned in each relevant category and the student's rated position within the overall distribution (Le Hénaff et al., 2015). Another option is to distinguish the higher achievers by indicating where they fit into a predetermined top percentile group, such as 10%, while not disclosing their individual information to their classmates.

5.1.5. Customization of badges

Pedro et al. (2015) found customization through digital badge awards to be an important empowering element. However, it was found in this study that customization should be considered carefully, especially in view of the ID group's desire to design and award badges themselves. Therefore, customization based on capturing the personalization of badge design and badge-issuing, and the self-awarding of badges and peer-issued badges via an automated Learning Management System (LMS) platform, should be given due attention.

5.1.6. Communications and interactions

Sun and Wu (2016) reported that interactions in an FC had a positive effect on students' learning achievement, but their findings regarding teacher-student interactions revealed that learners in both the experimental (flipped classes) and control groups (conventional classrooms) primarily conversed with peers and teaching assistants, but had little interaction with the instructor. However, it was found in this study that, as learners endeavored to build their knowledge outside class, they learned how to articulate their opinions and reach out to their peers or instructor for advice. As with the Engaged Achiever group, the Interactive Designer group perceived that teacher-student and student-student relationships were built within a collaborative space. Therefore, gamification using digital badges was found to have the potential to encourage learners not only to interact socially with their peers, but also with their instructors via out-of-class activities.

5.2. Conclusion

Educators' recognition of students' diverse preferences and different learning styles has been reinforced by the findings of this study based on its three groups; consequently, the use of digital badges and partially flipped classrooms needs to be encouraged at individual levels for teachers to appreciate the contrasting and concurrent perspectives of students driven by different motivations. For instance, it is suggested by the findings of this study that most female students pay more attention to social connectedness, whereas the majority of males are more interested in information seeking and challenging; hence, gamification may be gender-specific. The preliminary results of the study, as demonstrated by the EA, SME and ID groups, show the positive effect of digital badges in motivating and energizing students to engage in an educational milieu based on the Self Determination Theory. Above all, implementing these measures across the board could contribute to enhancing the effectiveness of the gamified digital badges system.

However, some elements of gamification appear to be changing the fundamental concept that physical rewards motivate students in both formal and informal learning settings. As the ID group questioned the appropriateness of using digital leaderboards, educators should be wary of their negative impact in terms of being demotivators, together with badges, especially when they are both compulsory and publicly displayed. Nevertheless, although still in the early adoption stage, badges may open up new possibilities for credential and assessment purposes; indeed, it has been shown in this study that instructors could use them to set clear expectations, offer choices, give interactive assignments, and provide timely feedback on individual students' progress in a flipped classroom setting.

5.3. Limitations of the study

The preliminary findings of this study should be interpreted with caution due to the small sample size; however, researchers may use the Q-statement results as starting points for hypothesis-testing research because they shed light on both the opportunities and challenges of new credential and assessment methods in FC. They may also be the precursor of an innovative instructional strategy centered on increasing adult students' motivation and eagerness to learn. However, due to the nature of Q-methodology, the results have not been statistically proven, pending further investigation, therefore the three types of learning styles that emerged from this study may only be considered as impressionistic. Since the integrated badges in this study were designed and implemented on the Moodle Learning Management System, the results cannot be generalized across other learning platforms or enhanced learning environments, such as augmented reality (AR) or virtual reality (VR). Unlike the emphasis on reliability and validity in R-methodology, these factors are not applicable to Q-methodology; rather, the views of the participants are what really matter when assessing the delivery of valuable results.

5.4. Recommendations

The research for this study was based on the use of gamified out-of-class activities in a flipped classroom in a business-oriented university. More research is needed to examine the effects of gamification on both in-class and out-of-class activities in similar conditions, with a particular focus on both in-class activities, based on the Self Determination Theory, and the value of using digital badges to foster gamification-centered positive learning outcomes. Learning styles should also be explored and the competitive context of digital leaderboards should be examined in depth for a better understanding of the social comparison effect on learning. Despite the promising findings of this study, it remains unclear whether leader badges, as described by the Factor C participants, have the same positive learning effect on low performers.

Additional proof is required by applying considerable effort to a discrete methodology to supplement the experts' opinions of the learning preferences identified in this study. Further investigation is also needed to determine if gender differences can affect students' perception of various gamification elements. A similar investigation should be conducted with a larger sample, different course, and various levels of education, for a more multi-faceted analysis of students' opinions and learning outcomes. Above all, future researchers should explore how academic institutions utilize digital badges with a view to benefitting all students' careers, as well as helping to fill a competency gap.

Acknowledgement

This research was funded by the National Science and Technology Council and the Ministry of Education, Taiwan, R.O.C. under Grant no. MOST 108-2511-H-216-003 and 108-Z27-MO-004.

References

Aldosari, M. A., Aljabaa, A. H., Al-Sehaibany, F. S., & Albarakati, S. F. (2018). Learning style preferences of dental students at a single institution in Riyadh, Saudi Arabia, evaluated using the VARK questionnaire. *Advances in Medical Education and Practice, 9*, 179-186. https://doi.org/10.2147/amep.s157686

Bhagat, K. K., Chang, C. N., & Chang, C. Y. (2016). The Impact of the flipped classroom on mathematics concept learning in high school. *Educational Technology & Society, 19*(3), 134-142.

Brown, S. R. (1993). A Primer on Q methodology. *Operant Subjectivity, 16*(3/4), 91-138. https://doi.org/10.22488/okstate.93.100504

Buckley, P., & Doyle, E. (2017). Individualising gamification: An Investigation of the impact of learning styles and personality traits on the efficacy of gamification using a prediction market. *Computers & Education, 106*, 43-55. https://doi.org/10.1016/j.compedu.2016.11.009

Cabi, E. (2018). The Impact of the flipped classroom model on students' academic achievement. *The International Review of Research in Open and Distributed Learning, 19*(3), 202-221. https://doi.org/10.19173/irrodl.v19i3.3482

Cela, K., Sicilia, M.-Á., & Sánchez-Alonso, S. (2015). Influence of learning styles on social structures in online learning environments. *British Journal of Educational Technology, 47*(6), 1065-1082. https://doi.org/10.1111/bjet.12267

Chen, L., Chen, T. L., & Chen, N. S. (2015). Students' perspectives of using cooperative learning in a flipped statistics classroom. *Australasian Journal of Educational Technology, 31*(6), 621-640. https://doi.org/10.14742/ajet.1876

Chen, T. L., & Chen, L. (2018). Utilizing Wikis and a LINE messaging app in flipped classrooms. *Eurasia Journal of Mathematics, Science and Technology Education, 14*(3), 1063-1074. https://doi.org/10.12973/ejmste/81548

Chen, Y., Wang, Y., Kinshuk, & Chen, N. S. (2014). Is FLIP enough? Or should we use the FLIPPED model instead? *Computers & Education, 79*, 16-27. https://doi.org/10.1016/j.compedu.2014.07.004

Chng, S., Chang, C., Mosquera, K., & Leong, W. Y. (2022). Living in a silver zone: Residents' perceptions of area-wide traffic calming measures in Singapore. *Transportation Research Interdisciplinary Perspectives, 16*, 100710. https://doi.org/10.1016/j.trip.2022.100710

Curry, L. (1987). *Integrating concepts of cognitive or learning style: A Review with attention to psychometric standards.* Canadian College of Health Service Executives.

Deci, E. L., & Ryan, R. M. (1985). The General causality orientations scale: Self-determination in personality. *Journal of Research in Personality, 19*(2), 109-134. https://doi.org/10.1016/0092-6566(85)90023-6

Deci, E. L., & Ryan, R. M. (2002). *Handbook of self-determination research.* University of Rochester Press.

DeLozier, S. J., & Rhodes, M. G. (2016). Flipped classrooms: A Review of key ideas and recommendations for practice. *Educational Psychology Review, 29*(1), 141-151. https://doi.org/10.1007/s10648-015-9356-9

de-Marcos, L., Garcia-Lopez, E., & Garcia-Cabot, A. (2016). On the effectiveness of game-like and social approaches in learning: Comparing educational gaming, gamification & social networking. *Computers & Education, 95*, 99-113. https://doi.org/10.1016/j.compedu.2015.12.008

Denden, M., Tlili, A., Essalmi, F., Jemni, M., Chen, N.-S., & Burgos, D. (2021). Effects of gender and personality differences on students' perception of game design elements in educational gamification. *International Journal of Human-Computer Studies, 154*, 102674. https://doi.org/10.1016/j.ijhcs.2021.102674

Facey-Shaw, L., Specht, M., van Rosmalen, P., & Bartley-Bryan, J. (2020). Do badges affect intrinsic motivation in introductory programming students? *Simulation & Gaming, 51*(1), 33-54. https://doi.org/10.1177/1046878119884996

Felder, R. M., & Silverman, L. K. (1988). Learning and teaching styles in engineering education. *Engineering Education, 78*(7), 674-681.

Felder, R. M., & Soloman, B. A. (1997). *Index of learning styles questionnaire.* https://www.engr.ncsu.edu/learningstyles/ilsweb.html

Hagger, M. S., Hardcastle, S. J., Chater, A., Mallett, C., Pal, S., & Chatzisarantis, N. L. (2014). Autonomous and controlled motivational regulations for multiple health-related behaviors: Between- and within-participants analyses. *Health psychology and behavioral medicine, 2*(1), 565-601. https://doi.org/10.1080/21642850.2014.912945

Hall, R., Jensen, R., & McLean, D. (2013). Learning in the geoscience classroom: Q-methodology, learning styles, and individual preferences. *Journal of Geoscience Education, 61*(1), 120-128. https://doi.org/10.5408/12-301.1

Hamari, J., Shernoff, D. J., Rowe, E., Coller, B., Asbell-Clarke, J., & Edwards, T. (2016). Challenging games help students learn: An empirical study on engagement, flow and immersion in game-based learning. *Computers in Human Behavior, 54*, 170-179. https://doi.org/10.1016/j.chb.2015.07.045

Hassan, M. A., Habiba, U., Majeed, F., & Shoaib, M. (2019). Adaptive gamification in e-learning based on students' learning styles. *Interactive Learning Environments, 29*(4), 545-565. https://doi.org/10.1080/10494820.2019.1588745

Hew, K. F., Huang, B., Chu, K. W., & Chiu, D. K. W. (2016). Engaging Asian students through game mechanics: Findings from Two experiment studies. *Computers & Education, 92-93*, 221-236. https://doi.org/10.1016/j.compedu.2015.10.010

Hu, J., Peng, Y., Chen, X., & Yu, H. (2021). Differentiating the learning styles of college students in different disciplines in a college English blended learning setting. *PLOS ONE, 16*(5), e0251545. https://doi.org/10.1371/journal.pone.0251545

Huang, E. Y., Lin, S. W., & Huang, T. K. (2012). What type of learning style leads to online participation in the mixed-mode e-learning environment? A Study of software usage instruction. *Computers & Education, 58*(1), 338–349. https://doi.org/10.1016/j.compedu.2011.08.003

Koivisto, J., & Hamari, J. (2014). Demographic differences in perceived benefits from gamification. *Computers in Human Behavior, 35*, 179-188. https://doi.org/10.1016/j.chb.2014.03.007

Kuo, M. S., & Chuang, T. Y. (2016). How gamification motivates visits and engagement for online academic dissemination– An Empirical study. *Computers in Human Behavior, 5*5, 16-27. https://doi.org/10.1016/j.chb.2015.08.025

Le Hénaff, B., Michinov, N., Le Bohec, O., & Delaval, M. (2015). Social gaming is inSIDE: Impact of anonymity and group identity on performance in a team game-based learning environment. *Computers & Education, 82*, 84-95. https://doi.org/10.1016/j.compedu.2014.11.002

Liou, W.-K., Bhagat, K. K., & Chang, C.-Y. (2016). Beyond the flipped classroom: A Highly interactive cloud-classroom (HIC) embedded into basic materials science courses. *Journal of Science Education and Technology, 25*(3), 460-473. https://doi.org/10.1007/s10956-016-9606-8

Lo, C. K., & Hew, K. F. (2020). Developing a flipped learning approach to support student engagement: A Design-based research of secondary school mathematics teaching. *Journal of Computer Assisted Learning, 37*(1), 142-157. https://doi.org/10.1111/jcal.12474

McKeown, B., & Thomas, D. B. (2013). *Q methodology.* Sage.

Mekler, E. D., Brühlmann, F., Tuch, A. N., & Opwis, K. (2017). Towards understanding the effects of individual gamification elements on intrinsic motivation and performance. *Computers in Human Behavior*, *71*, 525-534. https://doi.org/10.1016/j.chb.2015.08.048

Muilenburg, L. Y., & Berge, Z. L. (2016). *Digital badges in education: Trends, issues, and cases*. Routledge.

Pedro, L., Santos, C., Aresta, M., & Almeida, S. (2015). Peer-supported badge attribution in a Collaborative Learning Platform: The SAPO campus case. *Computers in Human Behavior*, *51*, 562-567. https://doi.org/10.1016/j.chb.2015.03.024

Riding, R., & Cheema, I. (1991). Cognitive styles—An Overview and integration. *Educational Psychology*, *11*(3-4), 193-215. https://doi.org/10.1080/0144341910110301

Seaborn, K., & Fels, D. I. (2015). Gamification in theory and action: A Survey. *International Journal of Human-Computer Studies*, *74*, 14-31. https://doi.org/10.1016/j.ijhcs.2014.09.006

Song, D., & Bonk, C. J. (2016). Motivational factors in self-directed informal learning from online learning resources. *Cogent Education*, *3*(1), 1205838. https://doi.org/10.1080/2331186x.2016.1205838

Stephenson, W. (1935). Correlating persons instead of tests. *Journal of Personality*, *4*(1), 17-24. https://doi.org/10.1111/j.1467-6494.1935.tb02022.x

Strelan, P., Osborn, A., & Palmer, E. (2020). Student satisfaction with courses and instructors in a flipped classroom: A Meta-analysis. *Journal of Computer Assisted Learning*, *36*(3), 295-314. https://doi.org/10.1111/jcal.12421

Subhash, S., & Cudney, E. A. (2018). Gamified learning in Higher Education: A Systematic review of the literature. *Computers in Human Behavior*, *87*, 192-206. https://doi.org/10.1016/j.chb.2018.05.028

Sun, J. C.-Y., & Hsieh, P.-H. (2018). Application of a gamified interactive response system to enhance the intrinsic and extrinsic motivation, student engagement, and attention of English learners. *Educational Technology & Society, 21*(3), 104-116.

Sun, J. C.-Y., & Wu, Y.-T. (2016). Analysis of learning achievement and teacher-student interactions in flipped and conventional classrooms. *The International Review of Research in Open and Distributed Learning, 17*(1), 79-99. https://doi.org/10.19173/irrodl.v17i1.2116

Thai, N. T. T., De Wever, B., & Valcke, M. (2017). The Impact of a flipped classroom design on learning performance in higher education: Looking for the best "blend" of lectures and guiding questions with feedback. *Computers & Education, 107*, 113-126. https://doi.org/10.1016/j.compedu.2017.01.003

Tinati, R., Luczak-Roesch, M., Simperl, E., & Hall, W. (2017). An Investigation of player motivations in Eyewire, a gamified citizen science project. *Computers in Human Behavior, 73*, 527-540. https://doi.org/10.1016/j.chb.2016.12.074

Urquiza-Fuentes, J. (2020). Increasing students' responsibility and learning outcomes using partial flipped classroom in a language processors course. *IEEE Access, 8*, 211211-211223. https://doi.org/10.1109/access.2020.3039628

Valaitis, R., Akhtar-Danesh, N., Eva, K., Levinson, A., & Wainman, B. (2007). Pragmatists, positive communicators, and shy enthusiasts: Three viewpoints on web conferencing in health sciences education. *Journal of Medical Internet Research, 9*(5), e39. https://doi.org/10.2196/jmir.9.5.e39

Valenta, A., & Wigger, U. (1997). Q-methodology: Definition and application in health care informatics. *Journal of The American Medical Informatics Association*, *4*(6), 501-510. https://doi.org/10.1136/jamia.1997.0040501

Waldrop, J. B., & Bowdon, M. A. (Eds.). (2016). *Best practices for flipping the college classroom*. Routledge.

Wang, Y., Liou, Y.C., Wang, Y., & Tang, T. (2004). Profiling the learning styles of students in cyber university. *Journal of Computer Information Systems, 45*, 113-122.

Watts, S., & Stenner, P. (2012). *Doing Q methodological research: Theory, method and interpretation.* Sage Publications.

Webler, T., Danielson, S., & Tuler, S. (2009). *Using Q method to reveal social perspectives in environmental research.* Social and Environmental Research Institute.

Ye, X.-D., Chang, Y.-H., & Lai, C.-L. (2018). An Interactive problem-posing guiding approach to bridging and facilitating pre- and in-class learning for flipped classrooms. *Interactive Learning Environments*, *27*(8), 1075-1092. https://doi.org/10.1080/10494820.2018.1495651

Vieira, C., Gómez, R. L., Gómez, M., Canu, M., & Duque, M. (2023). Implementing Unplugged CS and Use-Modify-Create to Develop Student Computational Thinking Skills: – A Nationwide Implementation in Colombia. *Educational Technology & Society*, *26*(3), 155-175. https://doi.org/10.30191/ETS.202307_26(3).0012

Implementing Unplugged CS and Use-Modify-Create to Develop Student Computational Thinking Skills: – A Nationwide Implementation in Colombia

Camilo Vieira[1*], Ricardo L. Gómez[2], Margarita Gómez[3], Michael Canu[4] and Mauricio Duque[3]

[1]Universidad del Norte, Colombia // [2]Universidad de Antioquia, Colombia // [3]ACCEFYN, Colombia // [4]Universidad El Bosque, Colombia // cvieira@uninorte.edu.co // rleon.gomez@udea.edu.co // mgomez@stem-academia.net // mcanu@unbosque.edu.co // ismaduque@stem-academia.net

[*]Corresponding author

(Submitted February 17, 2022; Revised July 18, 2022; Accepted November 20, 2022)

ABSTRACT: This paper describes the implementation and student learning outcomes of a nationwide professional development program for lower secondary and upper secondary school teachers to integrate computational thinking into the K-12 curriculum. Computational thinking comprises important concepts and skills that all students should develop to take an active role in a global society. However, teaching computational thinking is challenging. There are few teachers with the knowledge and skills to integrate computation into their courses. In this program, the participating teachers implemented a set of lesson plans that included both unplugged activities to scaffold student learning, and ‘plugged’ activities following a use-modify-create learning progression with the Micro:bit device to practice these skills. The study used a quasi-experimental design to compare students’ level of computational thinking between the program participants and a control group. The results suggest a positive effect of the learning activities on student computational thinking knowledge and skills as compared to the control group. This result persists after controlling for school context and student gender. This study provides an explicit approach to implementing these activities in the context of a developing country and assesses their effectiveness in a large-scale study.

Keywords: Computational thinking, Micro:bit, Use-modify-create, Unplugged, Teacher professional development

1. Introduction

Computational thinking (CT) is an important set of skills and practices that enable professionals from all disciplines to solve complex problems. Wing (2011) defined it as “the thought processes involved in formulating problems and their solutions so that the solutions are represented in a form that can be effectively carried out by an information-processing agent” (p. 1). Different definitions for computational thinking often involve concepts and practices such as abstraction, decomposition, automation, algorithmic thinking, and modeling and simulation (e.g., Grover, 2017; Katai, 2020; Noh & Lee, 2020). Such practices allow professionals to process large datasets, automate repetitive tasks, and understand and predict complex phenomena using, modifying, and creating computational models and simulations (Weintrop et al., 2016). Hence, computation has been denominated the third pillar of scientific discovery (Wing & Stanzione, 2016), together with theoretical and experimental approaches to inquiry.

The affordances of computing for professionals in all disciplines have generated a call from academics, governments, and international agencies to integrate computational thinking across all educational levels (Angeli et al., 2016; Barr & Stephenson, 2011; Katai, 2020; Lee et al., 2011; Royal Society, 2012). Some authors even suggest that developing student computational thinking is as important as learning math or reading and writing (Grover & Pea, 2013; Sanford & Naidu, 2016). However, teaching computational thinking may be more effective and inclusive when integrated into disciplinary contexts and real-world problems (Jocius et al., 2021; Mouza et al., 2020; Katai, 2020; Weintrop et al., 2016). Nevertheless, some degree of explicit direct instruction of basic concepts is required to avoid distraction and cognitive overload in students trying to make sense of a contextual project (Tricot, 2017).

Several countries, including the United States (Mouza et al., 2020), the United Kingdom (Curzon et al., 2014), and Australia (Yadav et al., 2017), have started to integrate computational thinking into their K-12 curricula. In many of these places, however, there is no formal training for teachers to incorporate computational thinking, which offers challenges for teaching it at the schools, and may stress teachers (Angeli et al., 2016; Caeli & Bundsgaard, 2020). Furthermore, most of what we know about how students learn computing is from developed countries and privileged students. More research is required to understand how students from different contexts (e.g., culture, language, technology access) learn about computational thinking and their challenges in this process.

This paper reports on the student outcomes after implementing a professional development program for lower secondary and upper secondary school teachers aimed at integrating computational thinking concepts and practices into the public school system in Colombia: "Programación para Niños y Niñas" (in English, Coding for Kids - CFK). Colombia, like many developing countries, faces four interrelated challenges. First, the education system is not producing enough people with CT skills to fill both current and projected industry demand. Second, women are significantly underrepresented in the STEM education pipeline and workforce. Third, many teachers have no preparation to integrate CT into their courses. Fourth, there is limited research on STEM teachers' pedagogical knowledge (PK) of CT and its effect on student learning.

Colombia is a country with significant achievement gaps between men and women and between urban and rural students. Colombia is the country with the most critical situation among all the countries belonging to the OECD: although Colombia follows the world trend of a better performance of women compared to men in Language skills, it has the smallest gap in favor of women. While in science and mathematics, the country presents the largest gap against women (OECD, 2020). The results of the 2018 PISA test also showed a systematic and large gap of almost 40 points between students in urban and rural areas, a value that compares with 24 points at the average level of OECD countries (ICFES, 2020). Bridging this divide between CT skills and those that students currently develop in Colombian schools requires innovative approaches to K-12 education (Barr et al., 2011; Mohaghegh & McCauley, 2016).

This paper aims to address these issues and assess the effects of a national-level professional development program on student learning. The research team designed a set of lesson plans that integrate computational thinking into their learning environments. Computer programming and CT skills represent a form of complex learning, where students need to consider many interacting elements at once (Mselle & Twaakyondo, 2012). Reducing extraneous cognitive loads and facilitating the development of schemata may contribute to effectively supporting student learning (Sweller et al., 2019). The lesson plans included a use-modify-create progression to scaffold student learning and reduce extraneous cognitive loads. While use-modify-create has been suggested as a promising approach to introduce CT skills (Lee et al., 2011), empirical evidence of how it works has just started to emerge (e.g., Franklin et al., 2020; Lytle et al., 2019), and little is known about how to operationalize its implementation. In this paper, we describe the design of Coding for Kids (CFK), its theoretical underpinnings, and its effects on students' CT skills. This study contributes to the body of knowledge of computational thinking education by providing and assessing an approach to integrate both unplugged learning activities and the use-modify-create learning progression in a developing country like Colombia. While we describe the teacher professional development program, our ultimate goal was to understand how it impacted student learning. To our knowledge, there is no such large-scale implementation of a professional development program that assesses student learning in a similar context. In this context, both teachers and students have limited knowledge about computing; the main language is Spanish (very few understand English), and the school infrastructure is limited. The guiding research question for this study is:

What is the effect of the Coding for Kids program on student computational thinking knowledge and skills?

2. Theoretical framework

2.1. Constructivism and cognitive load theory

The CFK program is based on the theoretical frameworks of both constructivism and cognitive load theory (CLT). Both of these frameworks informed the design of the lesson plans and can help explain potential differences in student performance on CT skills. From a Constructivist perspective, students build their knowledge starting from their prior knowledge while engaging in reflective practices (Ben-Ari, 1998; Jonassen et al., 1999). These student-centered learning environments promote changes in the existing cognitive structures and internal representations. Making connections to prior knowledge – including a mental model of what a computer is – and to relevant contexts for the student is key from this perspective (Ben-Ari, 1998).

Likewise, the Cognitive Load Theory (CLT) suggests a cognitive architecture that we use to process information and learn new concepts and skills (Sweller et al., 2019). The cognitive architecture includes a working memory (STM) and a long-term memory (LTM). The STM is limited in time and space, while the LTM is vast. There are different types of cognitive loads in the STM. The intrinsic load is given by the complexity of the concepts or skills to be learned, and can hardly be modified (Sweller et al., 2011). The extraneous load is not required for learning, and is often given by poor instructional designs or poor learning materials (e.g., trying to make sense out of the relationship between a poor figure and a text). The germane load, also called germane resources, is the positive load for learning. The germane resources help us make sense of the intrinsic load, by making connections to existing schemata (i.e., how knowledge is stored in the LTM). When we need to learn many interacting elements at once, we can experience a cognitive overload in our STM. We can only remember what we have in the STM if we make significant connections to the schemata stored in the LTM.

The CLT suggests several effects that may derive into effective pedagogical strategies to manage the cognitive loads in the STM. For instance, the worked-example effect describes how novice learners may reduce their cognitive loads by engaging in an active exploration of worked-examples instead of engaging in problem-solving from scratch. A worked-example is an expert's solution to a problem, which includes a problem statement, a step-by-step solution, and auxiliary representations of the problem and the solution (Atkinson et al., 2000). The worked-example effect suggests that by first exploring an expert's solution to a problem, novice learners start developing the required schemata that enable them to solve problems independently. At some point, the learning environment may include a fading approach by providing only partially solved worked-examples: the completion effect.

2.2. Implications for the lesson plans

We used two strategies in the design of the lesson plans to manage student cognitive loads. Start with unplugged learning activities to present basic concepts. Continue with computing activities using the Makecode editor and the Micro:bit device following the "use-modify-create" progression. The Micro:bit is an open-source programmable microcomputer to teach computer science concepts (Sentance et al., 2017). Physical computing devices such as the Micro:bit device have been explored to introduce computing concepts, and suggest positive outcomes on reducing student cognitive load while promoting creativity and motivation (Sentance et al., 2017). While these devices may be useful for computational thinking education, they require a pedagogical integration to support student learning. We designed seven lesson plans that teachers implemented in two separate 50-minute sessions. The first session is dedicated to an unplugged activity, while the second session is focused on the use-modify-create learning activity. Two additional lesson plans focused on assessment activities for teachers to monitor student learning.

This study explores the effect of integrating unplugged learning activities in computational thinking together with the "Use-Modify-Create" strategy on students' skills in Colombia. Hence, this study will contribute to understanding how these two approaches may support the development of CT skills, while providing a concrete approach to operationalize these approaches as a set of activities and scale it to a national level. This is particularly important in the context of a developing country such as Colombia, since most of the existing literature focuses on privileged students from developed countries.

2.2.1. Unplugged CS

Unplugged computer science is an approach to designing learning activities without needing a computing device. Computational thinking includes several skills and reasoning processes, many of which can be approached independently of the artifact: abstraction and generalization of patterns, information processing, symbolic representations and symbolic representation systems, algorithmic thinking, problem decomposition, iterative, recursive and parallel thinking, and conditional logic. Students can develop these skills without the need for a computing device. Instead, students may do "by hand" what a computer program would do in an automated way (Aranda & Ferguson, 2018; Faber et al., 2017). This approach ensures that the focus is on understanding and skill development and not on the artifact at an initial stage of the learning process (Bell et al., 2009). Starting with unplugged activities to learn these skills beyond a specific tool or programming language may help students to develop their problem-solving skills, and reflect on their own thinking (Caeli & Yadav, 2020).

Unplugged learning activities have been suggested as being effective both inside and outside the classroom at different levels to develop a fundamental understanding of concepts about algorithm design, and to change conceptions about the nature of computer science (Caeli & Yadav, 2020). This approach may reduce extraneous cognitive loads from elements that are inherent to some programming languages, like syntax or memory management. These activities also offer an efficient way to attract students to computer science concepts without the need for much time and resources (Bell et al., 2009, Brackmann et al., 2017). Despite the widespread use of unplugged activities, there is little empirical evidence in classroom settings (Bell & Vahrenhold, 2018).

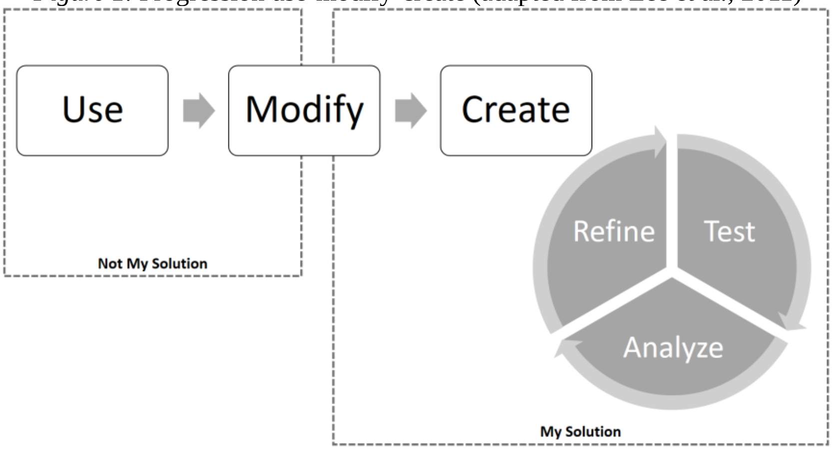
The development of most concepts and practices of computational thinking can be worked unplugged so that students can generalize these skills to other contexts that are not explicitly computer programming. However, researchers recommend using CS unplugged activities in conjunction with programming challenges (Bell & Vahrenhold, 2018). This should not be a discussion of whether CS unplugged is more powerful than using plugged programming lessons alone. CS unplugged activities may help students lower the cognitive loads associated with the device and programming language before actually applying this knowledge using a computer device.

Furthermore, integrating computational thinking at the k-12 level, especially in developing countries like Colombia, often brings challenges related to the accessibility to information and communication technologies for schools and students. This is especially the case for rural populations. At the beginning of this project, only 70% of the participating rural schools had working computing devices for their students (compared to around 90% for urban schools), and less than 50% of all participating schools had internet access. The program designed a series of unplugged activities to address this challenge and scaffold student learning.

2.2.2. Use-Modify-Create

The instructional principles of the worked-examples (Atkinson et al., 2000), explained by the CLT, support the Use-Modify-Create progression to scaffold student learning of complex concepts and skills. Computer programming is a complex skill to learn (Mselle & Twaakyondo, 2012), which may lead to cognitive overload. A novice learner needs to learn about algorithm design, the problem they are trying to solve, the programming language syntax and semantics, Boolean logic, and even how the computer works, all at the same time. The worked-example effect and the completion effect can be mapped to this progression as follows: the instructor provides a complete worked-example for students to explore (Use), students engage in the modification or completion of a worked-example (Modify), and then design and implement (Create) their own solutions iteratively (Figure 1).

Figure 1. Progression use-modify-create (adapted from Lee et al., 2011)



The progression Use-Modify-Create may help reduce such cognitive overload experienced by novice programmers. A recent study compared the use-modify-create progression to a sequence of activities where students always needed to create their own code (Lytle et al., 2019). In this study, the teachers in the create-only condition faced challenges supporting their students, and suggested that students needed additional scaffolding to explore and visualize a solution before they engaged in creating it. Conversely, teachers in the use-modify-create condition suggested that students were able to develop an understanding of how the program worked, and they connected everything once they completed the Create activity. Franklin and colleagues (2020) also found that this progression successfully scaffolds student learning, while promoting student agency. Finally, our own work has also shown that engaging students in the active exploration of examples may support the development of basic schemata for novice learners (Vieira et al., 2019).

Other approaches to support student learning based on the Cognitive Load Theory include sub-goal labeling (Morrison et al., 2015) and self-explaining (Vieira et al., 2017). However, the Use-Modify-Create provides a roadmap for students to increasingly develop their skills and agency (Franklin et al., 2020), moving from studying complete worked-examples, to faded solutions, to creating and refining their own solutions iteratively. The implications of this framework for this study are reflected in the learning design presented in the methods section.

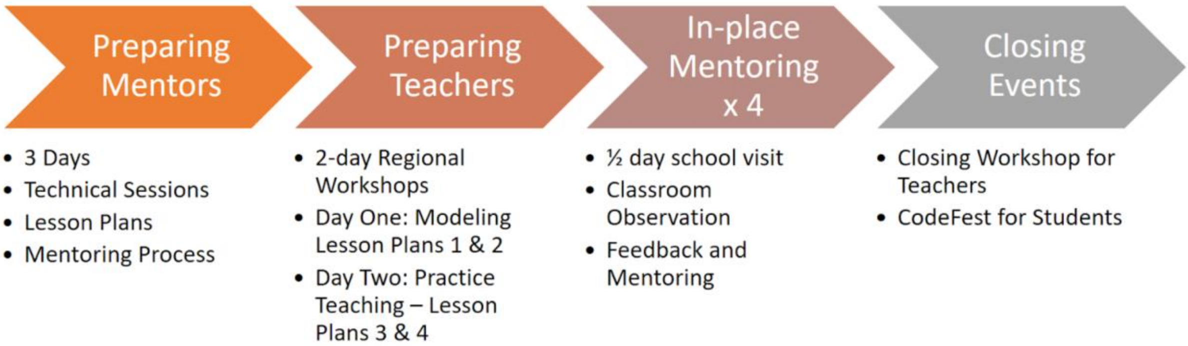
3. Methods

3.1. Program design

Despite the relevance of integrating computational thinking into the K-12 curricula, and the governmental efforts to make it happen, there is no agreement about what practices and skills computational thinking really entails (Curzon et al., 2019; Denning, 2017; Mouza et al., 2020). This makes it difficult to establish a universal set of learning outcomes to integrate into the school curricula (Mueller et al., 2017). For this project, the research team operationally defined computational thinking as a multidimensional set of complex skills, including computational problem-solving, computational modeling, abstraction, decomposition, algorithm design, programming and debugging, and solution validation. This definition integrates common concepts and skills that often appear in several authors' definitions (e.g., Denning, 2017; Grover, 2017; Katai, 2020; Weintrop et al., 2016).

Using this operational definition, the research team designed nine lesson plans to be implemented by participating teachers. While this paper focuses on student learning, the professional development program became our delivery mechanism for the learning activities. Figure 2 summarizes the implementation process of the learning materials. We first prepared a group of 12 mentors to lead the teacher training program. These 12 mentors implemented a two-day regional workshop for the teachers. The mentors modeled lesson plans #1 and #2 on the first day of the workshop, where teachers played the role of students. The participating teachers then prepared and taught lesson plans #3 and #4 for practice teaching on the second day of the workshop. The mentors also visited the teachers in their schools five times in a period of six months. Four of these monthly visits aimed at preparing the upcoming lesson plans, characterizing the participants' teaching practices and providing feedback and mentoring over the lesson plans. The goal of the fifth visit was to collect data on student performance.

Figure 2. Professional development program Coding for Kids



Each lesson plan comprises two 50-minute sessions, and follows the structure presented in Figure 3. The first session focused on unplugged activities, while the second session included plugged activities with the Micro:bit using the use-modify-create progression. The teachers start by discussing with students the learning outcomes for the activity and the required prior knowledge to complete it. This part allows students to manage their expectations, connect to their prior knowledge and, with a reflection at the end, promote metacognitive processes (Robins, 2019). Students work on the Micro:bit on the second session, following a use-modify-create progression. Students first explore an example, predicting the outcome of a program or explaining it to a partner (i.e., Use). Students then make specific changes to the sample program, to engage in a scaffolded problem-solving activity (i.e., Modify). Finally, students complete a challenge from scratch, engaging in problem-solving with the Micro:bit (i.e., Create), using what they learned from the example.

Figure 4 shows an example of this progression in lesson plan #3. The goal of this activity is to simulate an autonomous vehicle for scientific exploration inside a cavern. The proximity sensors are simulated using two buttons, while the Micro:bit display will show where the vehicle would go. Students first need to predict what the sample program is doing and run the program to validate the prediction (i.e., Use). This program identifies when a button is pressed and shows an arrow pointing toward the East. The next step is to Modify this program so that it shows all directions on the display depending on what buttons are pressed. Finally, students should Create a program using the compass sensor to guide the vehicle in one direction (e.g., North) following a black line. Both the Modify and the Create steps in the progression engage the student in an iterative process of implementing, testing, reflecting, and refining, often discussing them with a partner. The learning goals of this lesson plan included: Using Boolean input variables; Communicating instructions using the LED screen; Interpreting a sequence of instructions and a flow diagram to solve a problem such as a maze; Using logical operations; Using loops that are repeated until the task is finished. Students started with the unplugged activity, following and adjusting a flow diagram to move a toy around a sample maze to solve this challenge. These flow diagrams included loops and logical operations that students needed to program later, using the Micro:bit input variables (e.g., buttons) and LED display.

Figure 3. Structure of the lesson plans for Coding for Kids

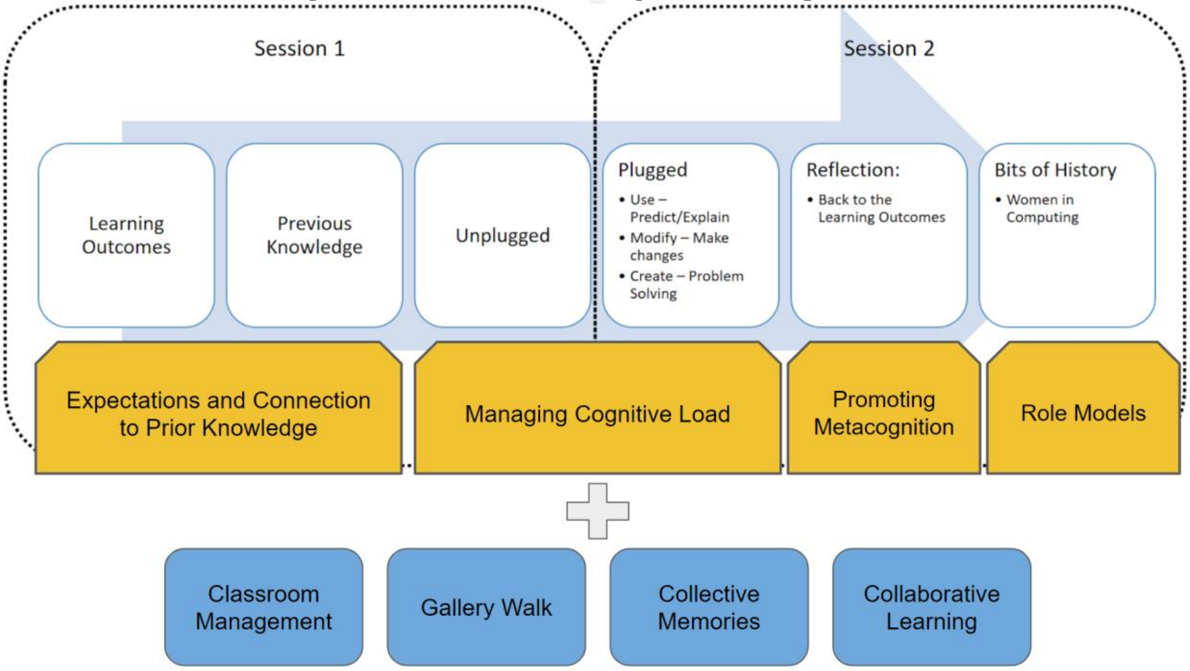
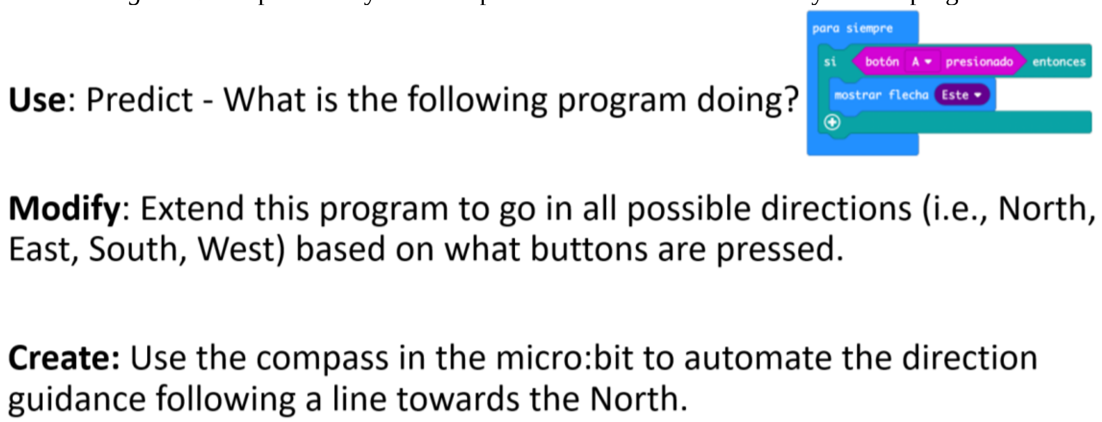


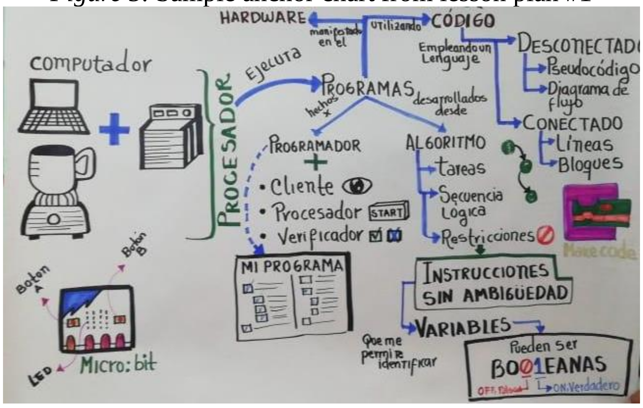
Figure 4. Sample activity of the implementation of the Use-Modify-Create progression



The unplugged activities and the use-modify-create progression help students manage the cognitive load by developing early schemata about the problem and the algorithm. At the end of the second session, the lesson plans include two additional elements: a reflection – connecting back to the learning outcomes – and a brief biography of a woman who worked in computing (Liben & Coyle 2014). The reflection aims at promoting student metacognitive processes such as monitoring and regulating. The goal of the brief biography was to present role models for female students, aimed at encouraging more female students to pursue STEM degrees. Sequencing activities within a learning design may contribute to student self-achievement and motivation, as they can see their progress through the outputs from each activity (Katai, 2020). We hypothesize that this sequence of activities in the lesson plan will result in student development of CT skills. Appendix A summarizes the learning outcomes for each of the lesson plans. Lesson plans five and nine represented assessment instruments for the teachers to monitor student learning, so they do not have specific learning outcomes for students to develop.

Besides the lesson plans, the program also included additional strategies to support the teaching and learning process. In Colombia, preliminary studies have demonstrated that only 65% of the time in the classroom is actually dedicated to teaching and learning, while the rest of the time is devoted to things like organizing students, social or recreational activities, or repeating instructions (MEN & The World Bank, 2012). To address this issue, the mentors modeled and made visible strategies such as setting norms and rules, promoting assertive communication between teachers and students, and promoting a growth mindset. Other strategies included the gallery walk, promoting collaborative learning, and the creation of anchor charts (e.g., Figure 5).

Figure 5. Sample anchor chart from lesson plan #1



3.2. Participants

A total of n = 282 lower secondary (grades 6-9) and upper secondary (grades 10-11) school teachers from 30 different municipalities in Colombia participated in the program. As part of the training, participants were required to implement the lesson plans in at least one group in lower secondary school or in upper secondary school, but some of them decided to use the activities for several groups. This resulted in approximately n = 20,000 students being included as the ultimate beneficiaries of the program.

The mentors scheduled their last visit to apply a computational thinking test in one group at each school. If there were multiple groups at different levels, the mentors chose an intermediate level (e.g., if the teachers implemented the activities in 6th, 7th, and 8th grade, the mentors would apply it for 7th grade). In total, 4077 lower secondary and upper secondary school students comprised the treatment group that completed a performance test measuring student CT skills.

The effect of the program on students' CT skills was assessed by means of a quasi-experimental posttest only design with a comparison group. Given the variety of schools, contexts, grade levels, and local curricula, students in the control group were chosen based on the following criteria: (1) the student's grade level, (2) that their teachers had not participated in the program, (3) that the control schools were located in the same municipalities as the participating schools, and (4) that they were not using a use-modify-create progression strategy in the classroom, though they were included if they were using a different computational thinking strategy or no strategy at all. Because there are no standard computer science or computational thinking curricula for Colombian public schools, the purpose of this quasi-experimental study is to determine whether our pedagogical approach, on average, supports student learning better than existing approaches or no approach at all. The final sample for the control group included n = 4898 seventh and eighth-graders. The demographic characteristics of the control group are shown in Table 1.

Table 1. Characteristics of the participating students

		Rural	Urban	Total
Treatment	Female	281	1694	1975
	Male	344	1758	2102
Control	Female	1062	1329	2391
	Male	1075	1432	2507
Total		2762	6213	8975

3.3. Data collection

The students were assessed using a multiple-choice test with 12 items, measuring skills such as abstraction, algorithmic thinking, automation, debugging, and solution validation. While there is recent work designing CT assessment instruments (e.g., Román-González et al., 2017; Werner et al., 2012), many of them are specific for a programming language, for an age group, or for a specific set of concepts, do not have a validated version for Spanish-speaking students, or are not available for open use. For this project, it was important to have programming language-independent questions in Spanish, well-aligned with the program goals, and be able to validate them locally. Our recent work showed that students in this context struggled to transfer into text-based programming languages, which mostly use English keywords (Espinal et al., 2022). The questions did not include the programming language MakeCode from Micro:bit since only the students in the treatment group were exposed to such language. Most of these questions were adapted from multiple sources (Bebras Computing Challenge, 2019; Grover et al., 2015; Tan & Venables, 2010) and included tasks such as explaining the goal of a program, tracking variables, suggesting a sequence of steps, and fixing bugs from simple flow-diagrams, pseudocode, and block-based programs. Table 2 shows the format, goal, and concept that each question assessed. This table also shows the CT skills for each item according to our operational definition, and the cognitive level of the item based on the number of correct responses. Appendix B includes all the items from this instrument.

The instrument was piloted in a sample of *n* = 171 students from non-participating schools using a cognitive lab approach (Willis et al., 1991). During a cognitive lab, participants are asked to complete a task and verbalize or "think aloud" the mental processes and thoughts occurring while solving the task. Students in the pilot sample did not participate in the intervention, but shared similar individual and educational characteristics of those participating in the study, including school grade level, age, and school characteristics. The discrimination and difficulty of each item were evaluated via item analysis (Wright, 2008). Six items with a probability of less than 20% being answered correctly and a low discrimination index were revised as a result of this analysis.

Table 2. Format, goal, and concept of the questions in the multiple-choice test

Item	Format	Goal	Concept	CT Skills	Cognitive Level
1	Blocks	Debugging	Loops	Algorithm Design, Decomposition, Abstraction, and Solution Validation	Easy
2	PseudoCode	Tracing	Conditionals	Algorithm Design; Programming and Debugging; Solution Validation	Medium
3	Descriptive PseudoCode	Debugging	Loops	Computational Problem Solving; Algorithm Design; Solution Validation	Medium
4	Descriptive PseudoCode	Debugging	Sequences	Computational Problem Solving; Algorithm Design	Difficult
5	Flow Diagrams	Explaining	Conditionals	Algorithm Design	Medium
6	PseudoCode	Debugging	Conditionals	Programming and Debugging; Solution Validation; Decomposition	Easy
7	PseudoCode	Tracing	Loops and Variables	Abstraction; Algorithm Design; Programming and Debugging; Solution Validation	Difficult
8	PseudoCode	Explaining	Loops, Conditionals and Variables	Abstraction; Algorithm Design; Decomposition; Programming and Debugging; Solution Validation	Difficult
9	Flow Diagrams	Explaining	Loops and Variables	Computational Modeling; Abstraction; Algorithm Design	Medium
10	Flow Diagrams-	Debugging	Loops	Computational Problem Solving; Algorithm Design; Solution Validation	Difficult
11	PseudoCode	Explaining	Loops, Variables, and Input/Output	Abstraction; Programming and Debugging	Medium
12	Flow Diagrams	Debugging	Variables and Conditionals	Computational Problem Solving; Computational Modeling; Decomposition	Difficult

The content validity of the instrument was evaluated with the data from cognitive labs with students and with face validity with two professional experts that contributed to the design of the lesson plans. This process allowed the researchers to improve the instrument. Specifically, one of the items did not include a correct answer, and there was an indentation problem with one of the pseudocode items. The researchers fixed these two issues in the instrument, and used it to assess student CT skills in the treatment and control groups towards the end of the academic year.

3.4. Data analysis

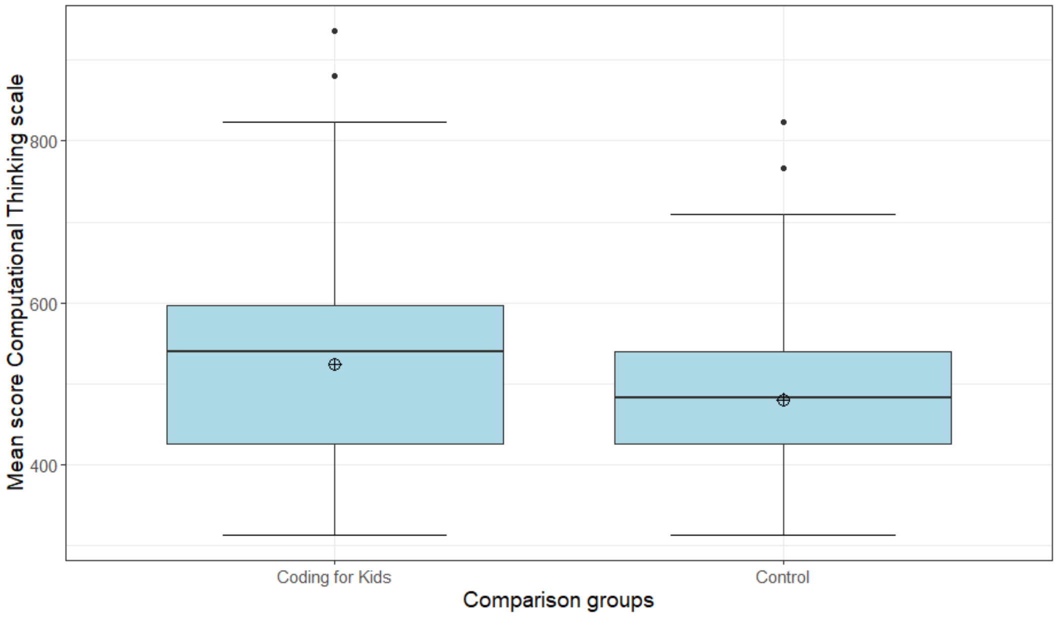
Students' responses were used to create a computational thinking scale (CTS) with a mean of 500 and a standard deviation of 100 (Streiner et al., 2015). For the construction of the scale, we first calculated the number of correct answers for each student and normalized the total of correct answers. By standardizing the scores, it is possible to compare the CT skills of the treatment group to the control group, since we can assess the position of each of the students above or below the average.

Using the CTS scores, we conducted a Welch's *t*-test to identify whether the differences in CT skills between the treatment and the control groups were statistically significant. Because the treatment and control groups have unequal sample sizes, Welch's *t*-test is preferred over the two-sample *t*-test statistic since it is more robust to violations of the assumption of variance homogeneity (Delacre et al., 2017). The Cohen's d effect size was also computed to identify how strong the effect of the program is. We used the scale suggested by Rubin (2012) to interpret Cohen's d effect size as follows: (1) Weak effect size: $|d| < 0.2$; (2) Weak to moderate: $0.2 < |d| < 0.4$; (3) Moderate: $0.40 < |d| < 0.65$; (4) Moderate to strong: $0.65 < |d| < 0.8$; (5) Strong: $0.8 < |d|$. We also used a one-way ANCOVA to see if the participants' levels of computational thinking differed significantly based on their school context (urban vs. rural) after controlling for gender. The data were examined to ensure that they met the ANCOVA assumptions. Skewness, kurtosis, and normality tests, as well as inspections of the histogram and normal Q-Q plot for computational thinking, all indicated that it was normally distributed.

4. Results

This section compared students' CT skills between the treatment and the control groups. The results show an average score for the treatment group of 525 (SD = 102), while the average score for the control group is 480 (SD = 93.9, see Figure 6). The Welch's t-test showed that the difference was statistically significant ($t(8389) = 21.4$, $p < .05$), with a moderate Cohen's d effect size of 0.45. The meaning of such effect size is that 67.4% of the participants in the treatment group performed better than the control group, and there is a 62.5% chance that a person randomly drawn from the treatment group would have a better performance than a person randomly drawn from the control group (Cumming & Calin-Jageman, 2017).

Figure 4. Coding for kids (treatment) vs. control group performance on the computational thinking scale



The one-way ANCOVA to control for gender ($F(1, 8967) = 11.418$; $p < .05$) and for school context ($F(1, 8967) = 119.189$; $p < .05$) are also statistically significant (see Figure 7 and Figure 8). Students in the treatment group showed better performance compared to students in the control group, both in urban and rural schools, and both male and female students. The mean difference between students enrolled in urban schools from the treatment and control groups is 31 points, while the difference is 58 points for rural schools.

When we analyzed the CT skills by grade level, both lower and upper secondary students in the treatment group scored considerably higher than their counterparts in the control group ($F(1, 9125) = 5.451$, $p < .05$). The mean difference between the treatment and control groups for lower secondary students is 34 points, while the difference for upper secondary students is 86 points (Figure 9).

Figure 5. Comparison of the distribution of student performance by context between coding for kids (treatment) and control groups

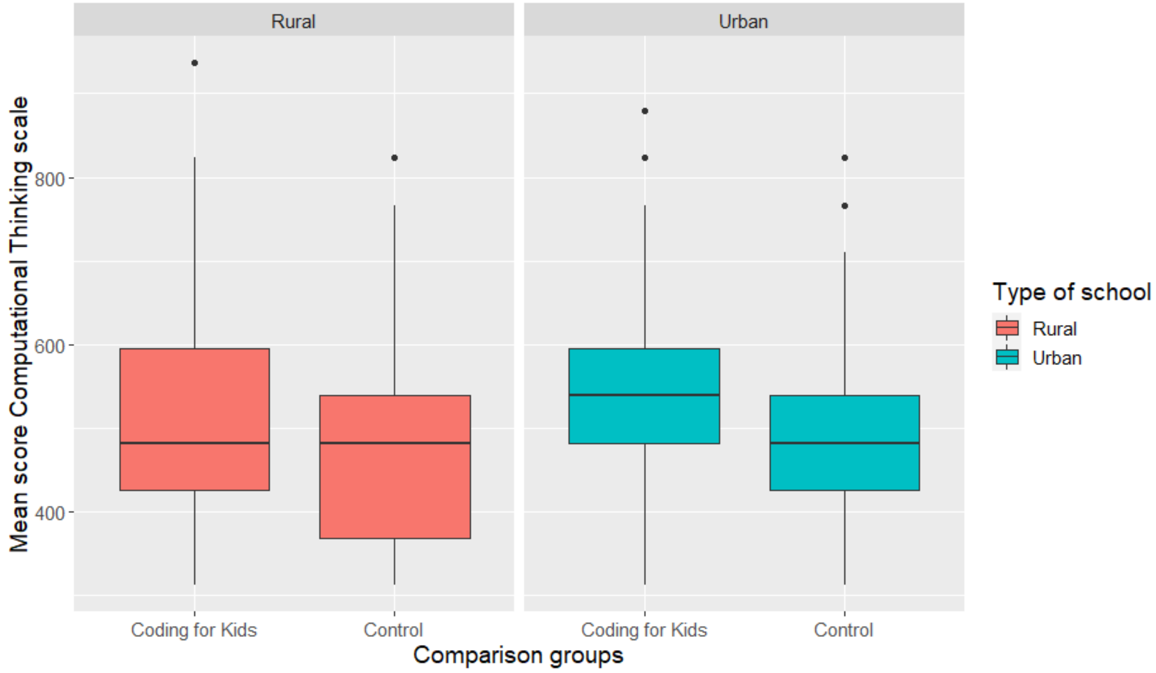


Figure 6. Comparison of the distribution of student performance by gender between Coding For Kids (Treatment) and control groups

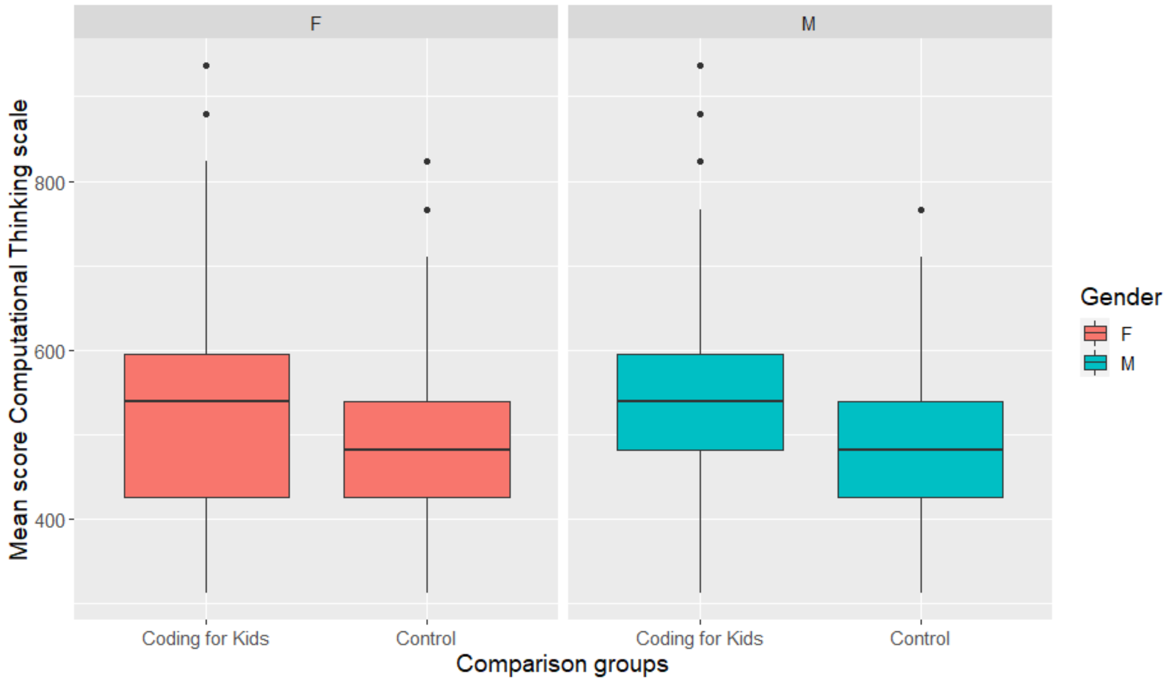
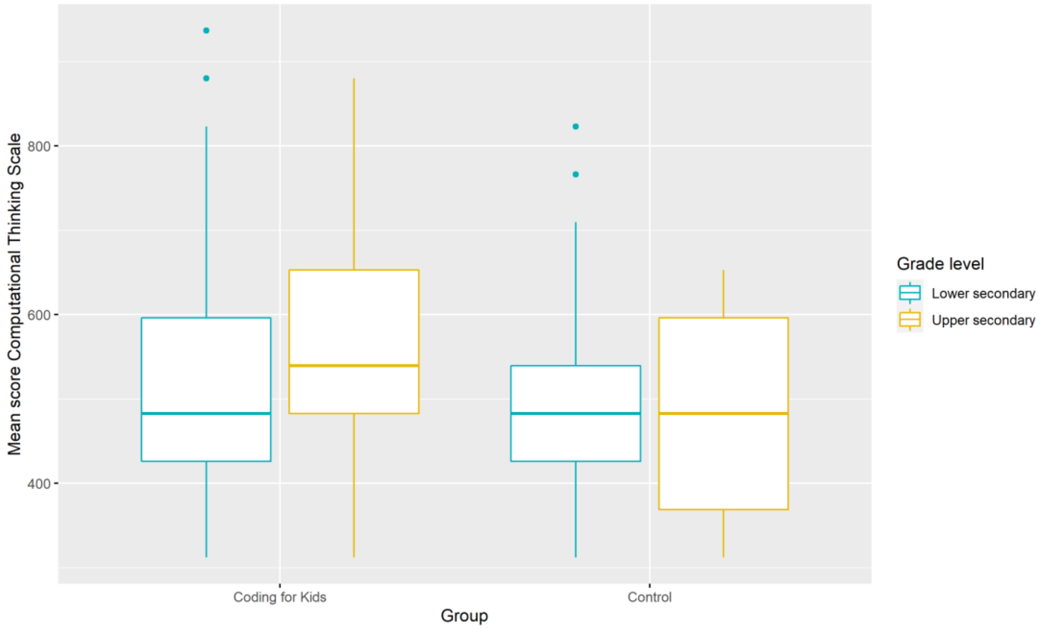


Figure 7. Comparison of student performance by grade levels between coding for kids (treatment) and control groups



5. Discussion

This study explored the *effect of the Coding for Kids project on student computational thinking knowledge and skills.* The Coding for Kids project put into action a professional development program for K-12 teachers to implement a set of lesson plans integrating computational thinking into Colombian public schools. The results from this study showed that students participating in these activities developed basic computational thinking knowledge and skills. This effect is evident when we compare participating students to a control group. Given the large scale of this study, it would be unfeasible for a single person to implement all the learning activities. While we recognize that there might be variability among teaching practices that may influence students' learning differently, our study shows that, on average, the professional development program and the learning activities have a positive effect on student learning. Although there are no pre-test data available for the experimental group, both the experimental and control groups were drawn at random from the same population. Given the large sample size, it is likely that both groups would have obtained similar results on a pre-test. On the basis of this assumption, it is reasonable to infer that the intervention significantly contributed to the higher posttest scores of the treatment group. The results were also supported by a Bayesian Mann-Whitney *U* test on independent samples (van Doorn et al., 2020), which provided strong support for the impact of the intervention. We observed a Bayes Factor (BF_{10}) of 8.906e+16, indicating a strong likelihood (8.906e+16 times in favor) of the alternative hypothesis relative to the null hypothesis. In other words, the alternative hypothesis that the students in the Coding for Kids group have higher levels of CT skills than students in the control, is strongly better at explaining the data than the null hypothesis of no difference between the two groups.

Before this project implementation, there was no specific approach to integrating CS or CT into the curriculum for Colombian public schools. Hence, the control group included any other possible approach that schools were using, including text-based programming languages, educational robotics, etc. Our results suggest that the Unplugged activities and the use-modify-create progression with a block-based programming language is, on average, more effective than other existing approaches. The effect persists when we control for context (i.e., urban and rural) and by gender. This effect is important because a large percentage of rural schools did not have access to computing devices or to an Internet connection to implement the plugged activities (i.e., approximately 70%). Thus, although the participating rural schools show greater performance variability than the urban ones, the unplugged learning activities still showed a significant effect for these schools with limited technological infrastructure. However, this effect does not mean that rural schools do not need access to the Internet or technological devices. Students from urban schools in the experimental group still showed a better performance than those from rural schools, suggesting that applying these concepts and skills using a computing device further supports student learning. Together, these results suggest that the use of unplugged learning activities and the use-modify-create progression (Lee et al., 2011; Lytle et al., 2019) may scaffold student learning of computational thinking knowledge and skills using the Micro:bit. Unplugged learning activities have become popular in recent years to support student learning without distractors and extraneous loads from the device and the programming language, promote collaborative learning, and attract more students to computer science (Bell et al., 2009, Brackmann et al., 2017). These activities help develop student understanding of basic concepts and have been suggested to be paired with plugged activities (Bell & Vahrenhold, 2018; Caeli & Yadav, 2020). We used these activities to introduce the algorithmic concepts before students engaged in programming the Micro:bit using MakeCode following the use-modify-create progression. Lee and colleagues (2011) proposed the use-modify-create progression as an approach to support student development of CT skills while maintaining the level of challenge for student development of schemata. A recent study showed promising results of the implementation of this progression based on teachers' reflections (Lytle et al., 2019) and student work (Franklin et al., 2020). In this study, we expand the evidence of the benefits of this progression by demonstrating a positive effect on student learning. While the results from the test show transfer to a certain degree (i.e., from the MakeCode language into text-based pseudocode), additional work is needed to identify whether these students are able to transfer their knowledge into text-based professional programming languages (Weintrop & Wilensky, 2017), and what challenges they face in this process. For example, a recent study showed that students are able to transfer between block-based programming languages, but they may struggle to transfer into programming languages such as Python (Espinal et al., 2022). The main difficulty seems to be that students do not understand the keywords in English when they learn using blocks in Spanish.

This study also advances our understanding of the challenges to integrating computational thinking in developing countries like Colombia. There are unique challenges in countries like Colombia for these activities, including the limited school infrastructure as well as the classroom management issues identified by previous studies (MEN & The World Bank, 2012). In a closing workshop at the end of the program, the participating mentors and teachers discussed several challenges they faced in this process. First, the limited technological infrastructure in the schools, particularly rural schools, made the program difficult to implement the plugged activities in such contexts. These teachers discussed issues such as "the lack of an internet connection in the school" and "the limited number of micro:bit devices" as challenges to the effective implementation of the lesson plans. Second, the program started in the middle of the academic year. This means that teachers already had a plan for the academic year, and they had to make adjustments, as they got involved in the professional development program. One of the participating teachers explained that the main challenge was "the time, because everything went so fast. Mostly, because we have several activities for this last part of the year." Yet another teacher highlighted that "the project is very good and it helps promote student interest, as well as developing student logical, computational, and creative thinking, but it needs to be introduced from the beginning of the academic year, so we can integrate it into the annual plan." Third, some classroom management issues persist, making classroom implementations inefficient. The participating teachers mentioned that even though they established norms, students do not always follow them: "the norms and instructions were not followed by some of the students." These challenges reflect that most of the teachers (70%) could only complete lesson plans one through five but did not have enough time to complete the rest of the lesson plans. The mentors, who conducted classroom observations and provided feedback to the participating teachers, ensured that the lesson plans were implemented following the design of the learning progression. Despite these challenges, this study showed how to operationalize unplugged activities and the use-modify-create progression into a set of lesson plans for a national professional development program, and identified a significant improvement on student CT skills when compared to a control group.

6. Conclusions, limitations, and next steps

This study explored the effect of the Coding for Kids program on developing students' CT skills. The program trained 282 k-12 teachers to integrate a set of lesson plans, aimed at developing student computational thinking in 30 different cities/areas in the country. Each lesson plan was designed for two sessions. The first session included a set of unplugged learning activities to prepare students for the second session. During the second session, students worked on the Micro:bit board following a use-modify-create progression designed to scaffold their learning process.

We created a computational thinking scale and compared participating students' performance to students in a control group (i.e., schools in similar contexts but that did not participate in the professional development program). Participating students outperformed students in a control group, both in urban and rural contexts. There are, however, some challenges involved in this process. We did not measure the effects of reducing the students' cognitive load to support student learning. Moreover, even if participating students performed better than non-participating students, the average number of correct responses in the test was low (i.e., 3.8 out of 12). The most difficult questions were those where students needed to transfer their learning about loops, a difficult concept in computer programming. There are several reasons to explain this phenomenon, including the limited infrastructure and project timing. Furthermore, the study only used a posttest to identify the possible effects of the program on student learning. Hence, any analysis of the significant differences that we found should consider these limitations. Future work will address these challenges by assessing cognitive loads and further controlling our experimental design by collecting a baseline for both experimental and control groups.

Acknowledgement

This work was funded by the Ministerio de TIC (Colombia), Computadores para Educar, and the British Council under the program Coding for Kids. The content is solely the responsibility of the authors and does not necessarily represent the official views of the supporting agencies.

References

Angeli, C., Voogt, J., Fluck, A., Webb, M., Cox, M., Malyn-Smith, J., & Zagami, J. (2016). A K-6 computational thinking curriculum framework: Implications for teacher knowledge. *Educational Technology & Society, 19*(3), 47-57.

Aranda, G., & Ferguson, J. P. (2018). Unplugged programming: The Future of teaching computational thinking? *Pedagogika, 68*(3), 279-292. https://doi.org/10.14712/23362189.2018.859

Atkinson, R. K., Derry, S. J., Renkl, A., & Wortham, D. (2000). Learning from examples: Instructional principles from the worked examples research. *Review of educational research, 70*(2), 181-214. https://doi.org/10.3102/00346543070002181

Barr, D., Harrison, J., & Conery, L. (2011). Computational thinking: A Digital age skill for everyone. *Learning & Leading with Technology, 38*(6), 20-23.

Barr, V., & Stephenson, C. (2011). Bringing computational thinking to K-12: What is involved and what is the role of the computer science education community? *ACM Inroads, 2*(1), 48-54. https://doi.org/10.1145/1929887.1929905

Bebras Computing Challenge. (2019). *Bebras computing challenge booklets*. http://www.bebras.uk/

Bell, T., Alexander, J., Freeman, I., & Grimley, M. (2009). Computer science unplugged: School students doing real computing without computers. *The New Zealand Journal of Applied Computing and Information Technology, 13*(1), 20-29.

Brackmann, C. P., Román-González, M., Robles, G., Moreno-León, J., Casali, A., & Barone, D. (2017). Development of computational thinking skills through unplugged activities in primary school. In *Proceedings of the 12th Workshop on Primary and Secondary Computing Education* (pp. 65-72). ACM. https://doi.org/10.1145/3137065.3137069

Bell, T., & Vahrenhold, J. (2018). CS unplugged—How is it used, and does it work? In *Adventures between lower bounds and higher altitudes* (pp. 497-521). Springer, Cham. https://doi.org/10.1007/978-3-319-98355-4_29

Ben-Ari, M. (1998). Constructivism in computer science education. *ACM SIGCSE Bulletin, 30*(1), 257-261. https://doi.org/10.1145/274790.274308

Caeli, E. N., & Bundsgaard, J. (2020). Computational thinking in compulsory education: A Survey study on initiatives and conceptions. *Educational Technology Research and Development, 68*(1), 551-573. https://doi.org/10.1007/s11423-019-09694-z

Caeli, E. N., & Yadav, A. (2020). Unplugged approaches to computational thinking: A Historical perspective. *TechTrends, 64*(1), 29-36. https://doi.org/10.1007/s11528-019-00410-5

Cumming, G., & Calin-Jageman, R. (2017). *Introduction to the new statistics: Estimation, open science, and beyond*. Routledge, Taylor & Francis Group. https://doi.org/10.4324/9781315708607

Curzon, P., Bell, T., Waite, J., & Dorling, M. (2019). Computational thinking. In *The Cambridge Handbook of Computing Education Research* (pp. 513-546). Cambridge University Press. https://doi.org/10.1017/9781108654555

Curzon, P., Dorling, M., Ng, T., Selby, C., & Woollard, J. (2014). Developing computational thinking in the classroom: A Framework. Computing at School. http://eprints.soton.ac.uk/id/eprint/369594

Delacre, M., Lakens, D., and Leys, C. (2017). Why Psychologists should by default use Welch's t-test instead of Student's t-test. *International Review of Social Psychology, 30*(1), 92–101. https://doi.org/10.5334/irsp.82

Denning, P. J. (2017). Remaining trouble spots with computational thinking. *Communications of the ACM, 60*(6), 33–39. https://doi.org/10.1145/2998438

Espinal, A., Vieira, C., & Guerrero-Bequis, V. (2022). Student ability and difficulties with transfer from a block-based programming language into other programming languages: A Case study in Colombia. *Computer Science Education,* 1-33. https://doi.org/10.1080/08993408.2022.2079867

Faber, H. H., Wierdsma, M. D., Doornbos, R. P., van der Ven, J. S., & de Vette, K. (2017). Teaching computational thinking to primary school students via unplugged programming lessons. *Journal of the European Teacher Education Network, 12*, 13-24.

Franklin, D., Coenraad, M., Palmer, J., Eatinger, D., Zipp, A., Anaya, M., White, M., Pham, H., Gökdemir, O., & Weintrop, D. (2020). An Analysis of use-modify-create pedagogical approach's success in balancing structure and student agency. In *Proceedings of the 2020 ACM Conference on International Computing Education Research* (pp. 14-24). https://doi.org/10.1145/3372782.3406256

Grover, S. (2017). Assessing algorithmic and computational thinking in K-12: Lessons from a middle school classroom. In *Emerging research, practice, and policy on computational thinking* (pp. 269-288). Springer, Cham. https://doi.org/10.1007/978-3-319-52691-1_17

Grover, S., Pea, R., & Cooper, S. (2015). Designing for deeper learning in a blended computer science course for middle school students. *Computer Science Education, 25*(2), 199-237. https://doi.org/10.1080/08993408.2015.1033142

Instituto Colombiano para el Fomento de la Educación Superior (ICFES). (2020). *Informe Nacional de Resultados para Colombia - PISA 2018* [National Results Report for Colombia - PISA 2018]. ICFES. https://www2.icfes.gov.co/documents/39286/1125661/Informe_nacional_resultados_PISA_2018.pdf/

Jonassen, D. H., Peck, K. L., & Wilson, B. G. (1999). *Learning with technology: A Constructivist perspective*. Merrill.

Jocius, R., O'Byrne, W. I., Albert, J., Joshi, D., Robinson, R., & Andrews, A. (2021). Infusing computational thinking into STEM teaching. *Educational Technology & Society, 24*(4), 166-179.

Katai, Z. (2020). Promoting computational thinking of both sciences-and humanities-oriented students: an instructional and motivational design perspective. *Educational Technology Research and Development*, *68*(5), 2239-2261. https://doi.org/10.1007/s11423-020-09766-5

Lee, I., Martin, F., Denner, J., Coulter, B., Allan, W., Erickson, J., Malyn-Smith, J., & Werner, L. (2011). Computational thinking for youth in practice. *ACM Inroads, 2*(1), 32-37. https://doi.org/10.1145/1929887.1929902

Lytle, N., Cateté, V., Boulden, D., Dong, Y., Houchins, J., Milliken, A., Isvik, A., Bounajim, D., Wiebe, E., & Barnes, T. (2019). Use, modify, create: Comparing computational thinking lesson progressions for STEM Classes. In *Proceedings of the 2019 ACM Conference on Innovation and Technology in Computer Science Education* (pp. 395-401). https://doi.org/10.1145/3304221.3319786

MEN, & The World Bank. (2012). Reporte del uso del tiempo en el aula: Evidencia para Colombia utilizando el método de observación de stallings. [Report on time use in the classroom: Evidence for Colombia using the stallings observation method].

Mohaghegh, D. M., & McCauley, M. (2016). Computational thinking: The Skill set of the 21st century. *International Journal of Computer Science and Information Technologies, 7*(3), 1524-1530.

Morrison, B. B., Margulieux, L. E., & Guzdial, M. (2015). Subgoals, context, and worked examples in learning computing problem solving. In *Proceedings of the eleventh annual international conference on international computing education research* (pp. 21-29). https://doi.org/10.1145/2787622.2787733

Mouza, C., Pan, Y. C., Yang, H., & Pollock, L. (2020). A Multiyear investigation of student computational thinking concepts, practices, and perspectives in an after-school computing program. *Journal of Educational Computing Research*, *58*(5), 1029-1056. https://doi.org/10.1177/0735633120905605

Mselle, L. J., & Twaakyondo, H. (2012). The Impact of Memory Transfer Language (MTL) on reducing misconceptions in teaching programming to novices. *International Journal of Machine Learning and Applications, 1*(1), 6. https://doi.org/10.4102/ijmla.v1i1.3

Mueller, J., Beckett, D., Hennessey, E., & Shodiev, H. (2017). Assessing computational thinking across the curriculum. In *Emerging research, practice, and policy on computational thinking* (pp. 251-267). Springer, Cham. https://doi.org/10.1007/978-3-319-52691-1_16

Noh, J., & Lee, J. (2020). Effects of robotics programming on the computational thinking and creativity of elementary school students. *Educational Technology Research and Development, 68*(1), 463-484. https://doi.org/10.1007/s11423-019-09708-w

Organisation for Economic Co-operation and Development (OECD). (2020). *Country note: Colombia*. https://www.oecd.org/pisa/publications/PISA2018_CN_COL_ESP.pdf

Robins, A. V. (2019). Novice programmers and introductory programming. In *The Cambridge Handbook of Computing Education Research* (pp. 327-376). Cambridge University Press. https://doi.org/10.1017/9781108654555

Royal Society. (2012). *Shut down or restart?: The Way forward for computing in UK schools*. Royal Society. https://royalsociety.org/~/media/education/computing-in-schools/2012-01-12-computing-in-schools.pdf

Román-González, M., Pérez-González, J. C., & Jiménez-Fernández, C. (2017). Which cognitive abilities underlie computational thinking? Criterion validity of the Computational Thinking Test. *Computers in human behavior*, *72*, 678-691. https://doi.org/10.1016/j.chb.2016.08.047

Rubin, A. (2012). *Statistics for evidence-based practice and evaluation*. Cengage Learning.

Sanford, J. F., & Naidu, J. T. (2016). Computational thinking concepts for grade school. *Contemporary Issues in Education Research*, *9*(1), 23-32. https://doi.org/10.19030/cier.v9i1.9547

Sentance, S., Waite, J., Yeomans, L., & MacLeod, E. (2017). Teaching with physical computing devices: The BBC micro: bit initiative. In *Proceedings of the 12th Workshop on Primary and Secondary Computing Education* (pp. 87-96). https://doi.org/10.1145/3137065.3137083

Streiner, D. L., Norman, G. R., y Cairney, J. (2015). *Health measurement scales: A Practical guide to their development and use* (Fifth edition). Oxford University Press. https://doi.org/10.1093/med/9780199685219.001.0001

Sweller, J., van Merriënboer, J. J., & Paas, F. (2019). Cognitive architecture and instructional design: 20 years later. *Educational Psychology Review*, 1-32. https://doi.org/10.1007/s10648-019-09465-5

Tan, G., & Venables, A. (2010). Wearing the assessment 'BRACElet'. *Journal of Information Technology Education: Innovations in Practice*, *9*(1), 25-34. https://www.learntechlib.org/p/111692/

Tricot, A., (2017). Les contraintes spécifiques des apprentissages scolaires [The Specific constraints of school learning]. *Psychologie & Education*, *2017-1*, 68-82. https://hal.science/hal-01628833

van Doorn, J., Ly, A., Marsman, M., and Wagenmakers, E.-J. (2020). Bayesian rank-based hypothesis testing for the rank sum test, the signed rank test, and Spearman's ρ. *Journal of Applied Statistics*, *47*(16), 2984–3006. https://doi.org/10.1080/02664763.2019.1709053

Vieira, C., Magana, A. J., Falk, M. L., & Garcia, R. E. (2017). Writing in-code comments to self-explain in computational science and engineering education. *ACM Transactions on Computing Education (TOCE)*, *17*(4), 1-21. https://doi.org/10.1145/3058751

Vieira, C., Magana, A. J., Roy, A., & Falk, M. L. (2019). Student explanations in the context of computational science and engineering education. *Cognition and Instruction*, *37*(2), 201-231. https://doi.org/10.1080/07370008.2018.1539738

Weintrop, D., Beheshti, E., Horn, M., Orton, K., Jona, K., Trouille, L., & Wilensky, U. (2016). Defining computational thinking for mathematics and science classrooms. *Journal of Science Education and Technology*, *25*(1), 127-147. https://doi.org/10.1007/s10956-015-9581-5

Weintrop, D., & Wilensky, U. (2017). Comparing block-based and text-based programming in high school computer science classrooms. *ACM Transactions on Computing Education (TOCE)*, *18*(1), 1-25. https://doi.org/10.1145/3089799

Werner, L., Denner, J., Campe, S., & Kawamoto, D. C. (2012). The Fairy performance assessment: Measuring computational thinking in middle school. In *Proceedings of the 43rd ACM technical symposium on computer science education* (pp. 215e220). http://dx.doi.org/10.1145/2157136.2157200

Willis, G. B., Royston, P., & Bercini, D. (1991). The Use of verbal report methods in the development and testing of survey questionnaires. *Applied Cognitive Psychology*, *5*(3), 251–267. https://doi.org/10.1002/acp.2350050307

Wing, J. (2011) Research notebook: Computational thinking - What and why? In *The Link Magazine* (pp. 20-23). Carneige Mellon

Wing, J. M., & Stanzione, D. (2016). Progress in computational thinking, and expanding the HPC community. *Communications of the ACM*. https://doi.org/10.1145/2933410

Wright, R. (2008). *Educational assessment: Tests and measurements in the age of accountability*. SAGE. https://dx.doi.org/10.4135/9781483329673

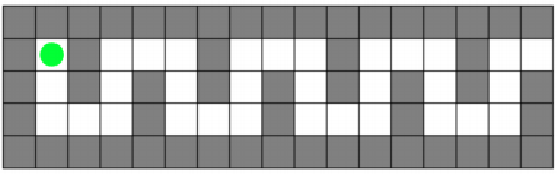
Yadav, A., Stephenson, C., & Hong, H. (2017). Computational thinking for teacher education. *Communications of the ACM*, *60*(4), 55-62. https://doi.org/10.1145/2994591

Appendix A – Learning outcomes of the lesson plans

Lesson plan	Learning goals
1	• Identify and write a set of steps and instructions to carry out a task. [Algorithm Design] • Simulate the execution of this set of instructions and steps. [Decomposition; Solution Validation] • Use the MakeCode editor of the Micro:bit to write a program and simulate. [Programming and Debugging] • Use inputs and outputs of the Micro:bit [Programming and Debugging] • Use Boolean variables. [Algorithm Design] • Describe what is a program, a programmer and a processor, an input and an output. [Programming and Debugging]
2	• Use loops to repeat a set of actions [Abstraction] • Recognize that a loop can be repeated indefinitely, a number of times or as long as a condition is met or not. [Abstraction; Algorithm Design] • Interpret and create flow diagrams. [Algorithm Design] • Load the code into the Micro: bit and verify the operation of the program [Programming and Debugging] • Use continuous input variables. [Programming and Debugging] • Show a variable like the temperature in the array of LEDs. [Programming and Debugging; Solution Validation]
3	• Use boolean input variables [Programming and Debugging] • Communicate instructions using the LED screen [Programming and Debugging] • Interpret a sequence of instructions and a flow diagram to solve a problem such as a maze. [Algorithm Design] • Use logical operations [Algorithm Design; Decompositon] • Use loops that are repeated until the task is finished. [Abstraction; Decomposition; Programming and Debugging]
4	• Define an internal variable that stores a numeric value [Programming and Debugging] • Perform operations with the values in these variables [Programming and Debugging]
5	• Test N/A
6	• Structure a problem situation. [Computational Problem Solving] • Identify specifications. [Computational Problem Solving] • Identify restrictions. [Computational Problem Solving] • - Design, build and test a prototype to evaluate some principles of the solution. [Programming and Debugging; Computational Problem Solving]
7	• Collect data from the environment (e.g., temperature) using the Micro:bit [Programming and Debugging] • Use conditionals to make decisions using the data collected from the environment [Algorithm Design; Decomposition] • Compute basic statistical measures such as the mean, and the min and max values [Computational Problem Solving] • Create a program that controls the functioning of the Micro:bit using the buttons [Programming and Debugging]
8	• Simulate natural events to predict possible outcomes [Computational Modeling] • Send and receive information between Micro:bit devices. [Programming and Debugging]
9	• Test N/A

Appendix B - Computational Thinking Skills Assessment Test

1. Help the green robot to exit the maze using one of the set of instructins below. Note: The number of times that the sequence repeats itself (3 or 4 times) will start counting after it executes the for the first time. For instance, if it says “3 times”, it will be executed 4 times in total.



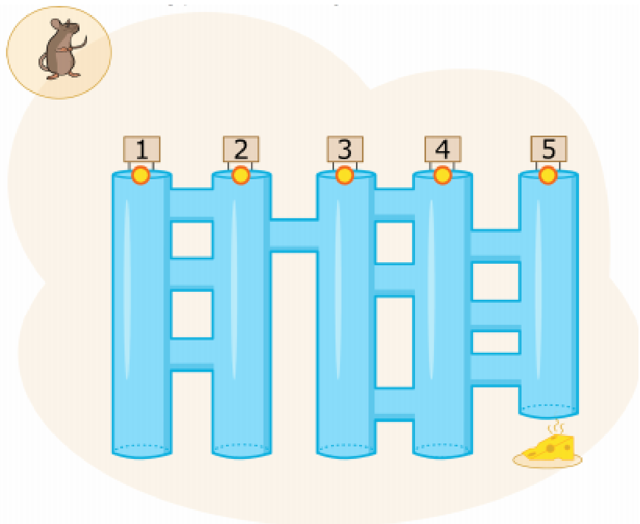
2. Consider the following program

```
if (a>b) then
    if(b<c) then
        print c
    else
        print b
else if (a<c) then
    print c
else
    print a
```

If a= 3, b= 8, and c=10, ¿what will this program print?

a. 3
b. 8
c. 10
d. 10 and 3

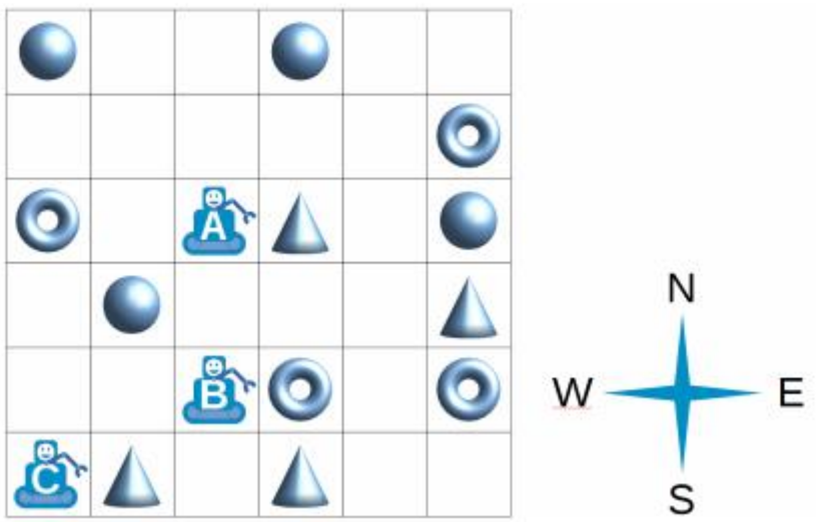
3. A mouse is at the entrance of a tube system. It wants to reach the cheese at the end of tube 5.The mouse always follows these command
(1) Go downwards until a crossing
(2) At the crossing, move through to the next vertical tube
(3) Go to command (1).
In which tube should the mouse start so that it reaches the cheese?



a. 1
b. 2
c. 3
d. 4
e. 5

4. In a warehouse, three robots always work as a team. When the team gets a direction instruction (N, S, E, W), all robots in the grid will move one square in that direction at the same time. After following a list of instructions, the robots all pick up the object found in their final square.

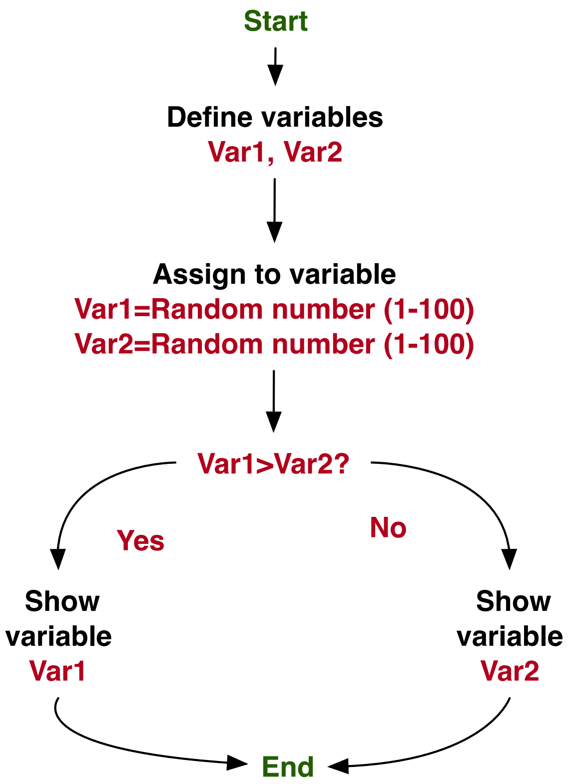
For example, if we give the list N, N, S, S, E to the team, then robot A will pick up a cone, robot B will pick up a ring, and robot C will pick up a cone



Which list of instructions can be sent to the robots so that the team picks up exactly a sphere, a cone, and a ring?

a. N, E, E, E
b. N, E, E, S, E
c. N, N, S, E, N
d. N, E, E, S, W

5. Consider the following flow diagram:



What is the goal of this algorithm?

a. Show two random numbers
b. Show the largest between two random numbers
c. Show the smallest between two random numbers
d. None of the above

6. Consider the following pseudocode, where a and b are variables:

```
a=3
b=5
if(a>2 and b<4) then
    print a
end
```

¿Will the *print a* instruction be executed?

7. Consider the following pseudocode. What will the program print at the end?

```
num=0
cont=1

while (cont<5) do
    num = num + cont
    cont = cont + 1

print num
```

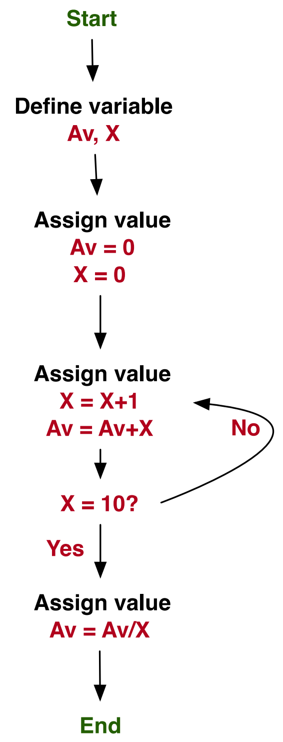
a. 15
b. 0
c. 1
d. 10

8. What is the purpose of the following program?

```
while (cont<10) do
    if cont even? then
        num = num + cont
    cont = cont + 1

print num
```

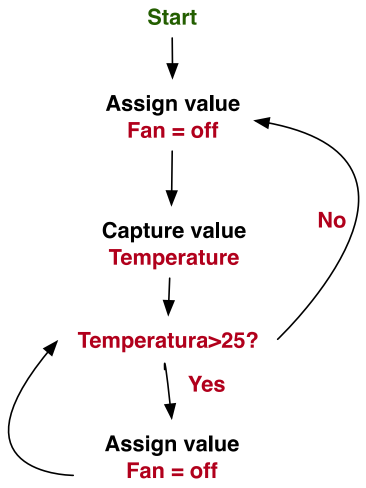
a. Creates a list including the even numbers between 1 and 10
b. Sums up all numbers from 1 to 10
c. Counts up to 10 if the numbers are even
d. Sums up all the even numbers between 1 and 9

9. What is the purpose of the following algorithm represented in a flow diagram?



a. Finds the number 10 to store it into the X variable.
b. Sums up all numbers between 1 and 10
c. Computes the average for the numbers from 1 to 10
d. Divides the Av value by all numbers from 1 to 10

10. Andrea created a flow diagram to design an algorithm that will allow her to automatically turn on the fan when her room is too hot. However, she is not sure this will work. What would you recommend?



a. The program does not work; the End (Fin) instruction is missing
b. The program does not work; she should not be turning the fan off
c. The program works correctly and she may implement it now to run in a processor
d. The program does not work; it should collect the temperature once more after turning the fan on

11. What is the purpose of the following program?

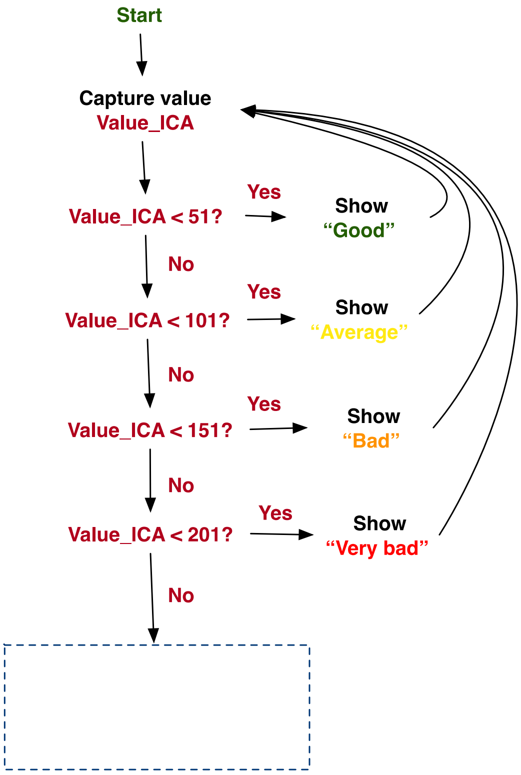
```
name = input('what is your name?')
cont=0

while cont<3:
    print('Hi, ')
    print(name)
    cont=cont+1
```

a. Asks for a name three times
b. Says hi, and print the name three times
c. Asks for a name, and says hi and prints the name twice
d. Asks for a name, and says hi and prints the name three times

12. Air pollution in the main cities is an important issue for the governors. Air pollution can be measured using the Air Quality Index (ICA, in Spanish).
The following table shows the ranges for a Good, Moderate, Bad, and Very Bad ICA

IDENTIFICADOR	ICA	CALIDAD DEL AIRE
	De 0 a 50	Buena
	De 51 a 100	Regular
	De 101 a 150	Mala
	De 151 a 200	Muy mala
	Más de 200	Extremadamente mala

Nick is in charge of creating a monitoring system to represent these values, and created the following flow diagram. However, he does not know how to finish it yet. What should he include in the dotted box?



a. Ask whether <u>*valorICA*</u> is smaller than 251, show "Extremadamente mala" and go back to *Capturar*
b. Show "Extremadamente mala" and go back to *Capturar*
c. Go back to *Capturar*
d. End

Lee, S. W.-Y., Liang, J.-C., Hsu, C.-Y., Chien, F. P., & Tsai, M.-J. (2023). Exploring Potential Factors to Students' Computational Thinking: Interactions between Gender and ICT-resource Differences in Taiwanese Junior High Schools. *Educational Technology & Society*, *26*(3), 176-189. https://doi.org/10.30191/ETS.202307_26(3).0013

Exploring Potential Factors to Students' Computational Thinking: Interactions between Gender and ICT-resource Differences in Taiwanese Junior High Schools

Silvia Wen-Yu Lee[1], Jyh-Chong Liang[2], Chung-Yuan Hsu[3*], Francis Pingfan Chien[2] and Meng-Jung Tsai[2]

[1]Graduate Institute of Information and Computer Education and Institute for Research Excellence in Learning Sciences, National Taiwan Normal University, Taipei, Taiwan // [2]Program of Learning Sciences and Institute for Research Excellence, National Taiwan Normal University, Taipei, Taiwan // [3]Department of Child Care, National Pingtung University of Science and Technology, Pingtung, Taiwan // swylee@ntnu.edu.tw // aljc@ntnu.edu.tw // jackohio@gmail.com // pingfan@ntnu.edu.tw // mjtsai99@ntnu.edu.tw // mjtsai99@gmail.com

[*]Corresponding author

(Submitted September 21, 2022; Revised December 26, 2022; Accepted January 16, 2023)

ABSTRACT: One of the major purposes of this study is to investigate the potential impact of gender and information and computer technology (ICT) resources on students' computational thinking (CT) competencies. To this end, the Computational Thinking Test for Junior High Students (CTT-JH) was developed and validated. Research participants included 437 junior high school students in Taiwan. The surveyed schools were categorized into *more* or *fewer* ICT resources. Then, discrimination analyses and Rasch modeling for item analyses and two-way ANOVA were conducted. Results showed that the final version of CTT-JH is of good item quality. Students in schools with more ICT resources had higher CT test mean scores regardless of gender. Nevertheless, at schools with limited resources, male students had significantly lower CT test mean scores than female students did. This study provides new insights into how gender and ICT resources can interact with and impact on students' CT competencies. It also provides a valid and reliable tool for assessing young adolescents' CT abilities.

Keywords: Computational thinking, Junior high school, Assessment, Non-programming, Domain-general CT

1. Introduction

Computational thinking (CT) can be regarded as one of the fundamental literacies of the 21st century for adapting to the future challenging society (Çoban & Korkmaz, 2021; Grover & Pea, 2018; Wing, 2006). The call for integrating CT into education has been gathering global attention during the last decade (Shute et al., 2017). Computational thinking (CT) refers to problem-solving skills (Wing, 2006) emphasizing conceptual development required to engage in formulating problems' solvable parts, abstracting key information, automating solutions through algorithmic thinking, debugging, and generalizing problem-solving processes (ISTE/ CSTA, 2011; Selby & Woollard, 2013; Shute et al., 2017). Individuals with CT skills are expected to execute the aforementioned skills to logically solve interdisciplinary and real-life problems (Araujo et al., 2019).

Various types of CT assessment instruments have been developed recently (Weintrop et al., 2021). While some instruments are to assess CT competencies based on programming and computing concepts, others are for assessing domain-specific or domain-general, non-programming problem solving competencies (Tang et al., 2020). We argued that among these different types of CT assessment, domain-general instruments that are congruent to the problem-solving perspective of CT and that can be assessed in non-computer science and even transdisciplinary learning context, require most attention from researchers and practitioners. As researchers have stressed the importance of fostering students' CT competencies at learning stages prior to college, there is a need for developing CT assessment for younger students as well (Li et al., 2020). Thus, in the current study, we developed a domain-general CT test for students at junior high school level and examined its psychometric properties. Furthermore, we are to explore the two potential factors that might impact students' CT competencies – gender and ICT-resources. The impact of gender on CT competencies remains inconclusive (e.g., Polat et al., 2021) and the relationships between ICT-resource abundance and CT competencies are by far under studied. In the following, a more comprehensive review of the different definitions of CT and the recent development of CT instruments will be introduced.

1.1. Computational thinking

A number of CT frameworks have been proposed in previous studies and these diverse frameworks imply that it is challenging to reach a consensus on CT operational definitions (Román-González et al., 2019; Shute et al., 2017; Voogt et al., 2015). Tang et al. (2020) categorized CT frameworks into two main aspects: (1) CT competencies related to programming and computing concepts, and (2) CT competencies needed for both domain-specific knowledge and general problem-solving skills. An example of the former is Brennan and Resnick's (2012) model consisting of computational concepts (sequences, conditionals, loops, etc.), practices (testing, debugging, reusing, etc.) and perspectives (viewing computation as a way of design and self-expression). An example of the latter is Shute's et al. (2017) CT model. Shute et al. (2017) synthesized 45 CT studies and proposed a competency-based CT model, not focusing on just one specific subject (e.g., coding) but approaching a problem-solving process in a systematic way. The model includes six main facets: decomposition (breaking a complicated problem into manageable parts), abstraction (identifying essential information), algorithms (logically developing solutions to a problem), debugging (finding and fixing errors), iteration (refining solutions), and generalization (transferring CT skills to other domains or situations). Similarly, Selby and Woollard (2013) reviewed CT studies and then proposed a CT model with abstraction, decomposition, algorithmic thinking, evaluations, and generalization. In sum, these frameworks can provide not only operational definitions for CT but also a foundation for CT assessments.

1.2. CT assessment

Assessments play a crucial role in determining successful integration of CT into educational contexts (Cutumisu et al., 2019; Poulakis & Politis, 2021; Tsai et al., 2022). With valid and reliable CT assessments, one can accurately evaluate students' CT development and understand the impact of the intervention (Eloy et al., 2022; Mueller et al., 2017). A majority of the assessment has been developed recently based on programming or computing concepts (i.e., the first type of instrument defined by Tang et al., 2020). For example, Román-González et al. (2018) developed Computational Thinking Test (CTt), a multiple-choice instrument for measuring learners' developmental level of CT based on fundamental programming concepts such as sequences, loops, and conditionals. Various instruments and assessment methods has been developed for measuring students' programming-based and computing-based CT. For instance, programming-based CT assessment also can be done through assessing students' programming artifacts or portfolio (Fields et al., 2021), through online puzzling games (Guenaga et al., 2021), or observing or logging students interactions (Metcalf et al., 2021). While some of the assessment utilized paper-based instruments designed for children in kindergarten in unplugged coding context (Clarke-Midura et al., 2021); others used computer automatic scoring for particular programming language, such as using Dr. Scratch for scoring Scratch-based programming artifacts (Moreno-León et al., 2015).

Nevertheless, Wing (2006) suggested that CT should not be limited to computer science or computer scientists, and further argued that CT involves computer concepts used by everyone to solve problems, manage their daily life, and interact with other people. The call for strengthening students' domain-general CT competencies has been receiving increasing attention; nevertheless, so far fewer instruments assessing domain-general CT skills and competencies are available (Tsai et al., 2021; Angeli & Giannakos, 2020; Kwon et al., 2021). Domain-general CT refers to "solving complex problems in daily life contexts" (Tsai et al., 2021, p. 2). In this sense, domain-general CT is even more important than domain-specific CT for developing future citizens' competencies for the 21st century. For instance, Tsai et al. (2022) has found that students' CT dispositions in problem-solving significantly predicted their domain-general CT competencies at elementary school level. A widely used domain-general CT assessment is the Bebras Challenge, a competition using real-life tasks to assess students' CT skills independent of previous programming experience (Dagienė & Stupuriene, 2016). The Bebras Challenge is hosted annually and internationally and more than 40 countries world-wide have participated. Example Bebras tasks can be seen at https://www.bebras.org/examples.html. Moreover, the domain-general CT instruments can be applied to various learning contexts and be utilized to examine students' CT competencies after different treatments. For instance, Chiazzese et al. (2019) measured the impacts of a robotics laboratory on the third and fourth graders' acquisition of CT competencies by using the Bebra tasks. The results showed that programming robots had a positive impact on students' acquisition of CT competencies.

Another important trend of recent research of CT assessment development is the attention to the quality of the research instruments and the scoring rubrics (Clarke-Midura et al., 2021). Researchers have raised the concerns of the lack of evaluating and reporting the validity and reliability of CT assessment in past publications (e.g., Tang et al., 2020). Using systemic methods for instrument development, such as evidence-centered design, and providing evidence of psychometric properties of instruments have been suggested by researchers when developing CT assessment (Basu et al., 2021; Clarke-Midura et al., 2021). Finally, there has been increasing attention of teaching CT for pre-college students (Weintrop et al., 2021). While recent development has shown a growth trend in CT research and CT measurement in elementary level (e.g., Basu et al., 2021; Metcalf et al., 2021; Polat et al., 2021; Tsarava et al., 2022;), the same growth has not been found at junior high school (or middle school) level. It is important to have domain-general CT instruments available at all levels for summative evaluation purposes and for monitoring students' learning progression. In the current study, we documented the process and evidence of validating a newly developed domain-general CT assessment for junior high school level (age 13-15).

1.3. Gender and digital divide

Additionally, in this study, we also aimed to explore the role of two factors in students' CT competencies–gender differences and the digital divide. Gender differences play a critical role in influencing students' CT development (Angeli & Valanides, 2020; Shute et al., 2017). Despite the fact that gender differences have been receiving growing attention recently, the findings from empirical studies seem to be inconclusive. Some studies have shown that males outperform females on CT tests at the secondary educational level (Guggemos, 2021; Tsai et al., 2022), and researchers have even reported that the higher the grade, the more intense the gender gap in CT performance (Román-González et al., 2017). Polat et al. (2021) implemented an intervention of visual programming, and found that male students tended to have better CT performance than that of girls (Polat et al., 2021). Nevertheless, in Durak and Saritepeci's (2018) study of secondary and high school students, they reported no significant relationships between gender and CT competencies.

Researchers have examined other factors, such as the type of activities, the time spent on task, or academic achievement in relation to gender differences in CT. For instance, utilizing educational robots to enhance students' CT, Angeli and Valanides (2020) found that male students benefited more from individualistic, kinesthetic, manipulative-based activities, whereas female students learned more from collaborative activities. While no statistically significant difference was found in students' CT competencies, Atmatzidou and Demetriadis (2016) found that female students required more time to reach the same CT level as males in educational robotics tasks. Furthermore, Lei et al. (2020) identified a stronger relation between CT and academic achievement among females than males in their meta-analysis research.

The digital divide is commonly defined as inequality in the use of information and communication technologies (Aydin, 2021; Light, 2009). Many researchers agree that unequal exposure to computers and advanced technology in general may impact students' interests in computer-based activities for learning and even hamper students' learning approaches and performance. Past studies have investigated the digital divide attributed to socio-economic status (SES) or the geographical location of schools. For instance, Hohlfeld et al. (2017) found that students in low-SES schools tended to use software for tutoring or practicing, while those from high-SES schools were inclined to use software more for researching, communication, and developing projects to demonstrate what they had learned. Moreover, Zhang (2014) utilized Google Trends and Web analytics to investigate middle and elementary school students' usage of the PhET website, one of the most well-known online science simulation resources. The results showed that students in high SES families were more interested in using PhET for learning sciences than their low-SES counterparts. In terms of geographical location, Kale et al. (2018) found that school rurality may influence teachers' own CT competence and their teaching of CT in classrooms. In other words, rural primary school teachers tended to have limited CT skills and felt that they were not ready to integrate CT into their teaching.

In the current study we examined the digital divide by using the schools' information and computer technology (ICT) resources as an indicator rather than school location. It is our observation that in Taiwan, school location does not necessarily contribute to the abundance or lack of ICT resources. In other words, rural schools or county-funded schools may have equal or more ICT resources than urban schools if they are enlisted as one of the ICT-schools or if the school is ambitious in getting more funding for ICT.

1.4. Purpose

Although past studies have shown the impact of digital divide on ICT competence or ICT attitudes, few studies, on the one hand, have examined its impact on students' CT competence or CT perceptions. As computational thinking is a 21st century skill in the current technological world, gaining more insights into how the digital divide influences students' CT development has become essential (Czerkawski & Lyman, 2015). On the other hand, it remains inconclusive under which conditions gender differences existed in CT performance. While gender and the digital divide are important issues in ICT literacy (e.g., Kim et al., 2021), one of the major purposes of this study is to investigate the potential impact of the gender and the ICT-resources, and their interactions on students' CT competencies.

To this end, it is important to have a valid, reliable, age-appropriate, and domain-general CT assessment instrument for researchers and teacher. A CT assessment tool, the Computational Thinking Test for Junior-High School Students (CTT-JH) was developed and validated in this study. In this study, the test items were adapted from the items from the Bebras Challenge and we revised the language, context, and presentation to make it suitable for students in Taiwan. The Bebras Challenge has been adopted internationally and has reached success in promoting computational thinking worldwide, however, only a few studies have examined its psychometric properties for research purposes. While some studies have used content analysis or success rate for analyzing item difficulty (Izu et al., 2017; van der Vegt, 2018), we suggested using Rasch modeling based on Item Response Theory (IRT) to provide more rigorous evidence regarding item quality. In sum, we posed the following research questions: (1) what are the validity and reliability of CTT-JH? (2) What are the effects of gender and ICT-resources on students' CT competencies?

2. Methods

2.1. Participants

Participants of the current study were 437 junior high school students in Taiwan, including 234 males (about 53.5%) and 203 females. Among the participants 105 were seventh grade students (about 24.0%), 162 eighth graders (about 37.1%), and 170 ninth graders (about 38.9%). The students were recruited from 16 intact classes in six junior high schools (two from a city, one from a county, two from a rural area, and one from a remote area) in the north and the center of Taiwan. To meet the ethical requirements, the participants were informed that their involvement in the study was voluntary and that their personal information would be treated confidentially. They were informed that they may withdraw from the study at any time. The students who agreed to participate then complete the CTT-JH items within an hour. The response rate was about 97.1%. School ICT resource information was collected from the school ICT administrator. All participants were assumed to have problem solving experience in their daily lives as well as in academic learning domains such as mathematics and science. They also had experience of using ICT for learning before participation.

The six schools were divided to more or fewer ICT resource groups based on the following three criteria : (1) the ratio of full-time ICT teacher to class, if the ratio for a school was greater than 0.1 then the school was coded as 1, otherwise was coded as 0 ; (2) the funding of Maker Education and Technology Center from the government, the schools with the findings were coded as 1, otherwise was coded as 0; and (3) the implementation of project-based ICT-integrated curriculum, the school that conducted the curriculum was coded as 1, otherwise was coded as 0. These criteria indicated the likelihood for students to be taught by full-time ICT teachers, the school involvement in maker education, and their implementation of ICT-integrated curriculum. Data were obtained from the Ministry of Education of Taiwan during 2019-2020. After the coding, we found two schools had ICT-teacher to class ratio more than 0.1; three schools received funding during 2019-2020 for maker centers; and two schools had projects for designing ICT-integrated curriculum. By summing up the three indices, each school obtained a total score that indicated the ICT resource of the school. If a school received a total score of 0, then the school was categorized into the fewer-resource group; otherwise it was categorized into the more-resource group. For detailed information of each school, please see Appendix. Finally, four schools were labeled as more-resource schools and two schools were labeled as fewer-resource schools.

2.2. Research instrument

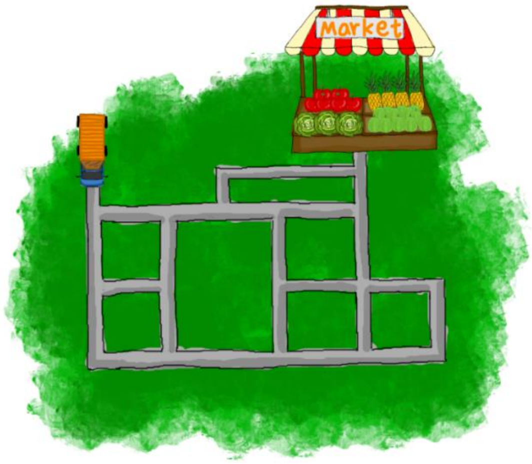
The Computational Thinking Test for Junior High students (CTT-JH) was a test developed to measure junior high school students' CT performance in this current study. Revised from the Bebras Challenge tasks, a pool of 15 items built up the initial version of the CTT-JH for assessing abstraction, decomposition, algorithmic thinking, evaluation, and generalization. We also referred to the CT framework in which these five dimensions respectively refer to the ability to abstract essential information, to break down complicated problems into manageable parts, to think procedurally as a sequence of steps to reach a solution, to decide the most appropriate solution to the problem, as well as to adapt and transfer solutions to other problems. Each item was designed to assess one or more CT dimensions simultaneously, and collectively CTT-JH is a multi-dimensional research instrument. All the items were redesigned or modified as solving problems in a daily-life farm-based context in

Taiwan. Two sample items are illustrated in Figure 1 and the complete test is available online at https://bit.ly/2022-CTT-JH.

Figure 1. Sample items of CTT-JH

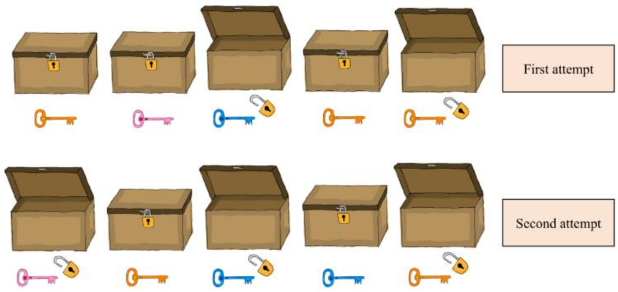
Q2. Selling goose eggs (easy)
A truck from Jack's Happy Farm is on the way to the market to sell goose eggs. The truck can drive in only 3 ways:

1. Left: Turn 90 degrees left
2. Right: Turn 90 degrees right
3. Forward: Go forward until you cannot go forward anymore



Question: Write a set of instructions (a program) that will get the truck to the market. You can do this by writing down the code numbers.

Q4. Hungry goose (difficult)
On Jack's Happy Farm, a hungry goose is trying to unlock five food boxes. Jack gives the goose 3 keys of different colors and says that these keys can open all the boxes. The results of the goose's first and second attempts are shown below.



Question: Which one is the correct order of the keys to open all the boxes?

1. Blue, Pink, Blue, Orange, Orange
2. Pink, Blue, Blue, Blue, Orange
3. Pink, Blue, Blue, Pink, Orange
4. Pink, Pink, Blue, Pink, Orange

2.3. Data analysis

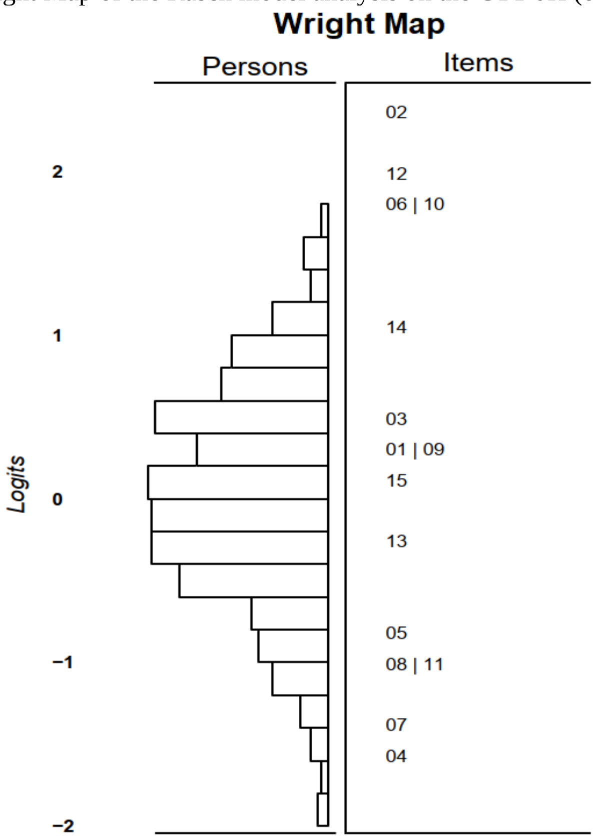
To ensure the content validity of the CTT-JH, the items were reviewed by the research team who had expertise in computer education, educational technology, and science education. Through meetings, consensus about the CT dimensions assessed by each CTT-JH item was developed among the experts. A list of items and its corresponding CT dimensions will be presented in the result section. To understand whether the CTT-JH test items are fitted and reliable measurement for junior high school students, we applied the Rasch model, a one-parameter logistic Item Response Theory (IRT) model for dichotomous items (Andrich & Marais, 2019; Mayer et al., 2014; Rasch, 1960) for data analysis. Test Analysis Modules (TAM) and the Wright Map packages in R software were used to estimate item difficulties and students' abilities on the same logit scale (Robitzsch et al., 2020; Irribarra & Freund, 2014). Finally, in order to examine whether there was any significant differences in the participants' CTT-JH test scores due to gender and teaching resources, a 2 x 2 ANOVA (gender x resource) was conducted.

3. Results

3.1. Psychometric properties of the CTT-JH items

Figure 2 illustrates the Wright Map of the Rasch model analysis of the participants' CTT-JH scores. It shows the distributions of the student's abilities (on the left side) and the distributions of the item difficulties (on the right side). The original 15 items were ordered from the most difficult (at the top, i.e., item 2) to the least difficult (at the bottom, i.e., item 4). The histograms of student's abilities show that each student solved the item with a probability of 50% and are plotted from most able (at the top) to least able (at the bottom).

Figure 2. The Wright Map of the Rasch model analysis on the CTT-JH (original 15 items)



After we fitted the Rasch model for the original 15 items, we examined the reliability of the whole test. The Weighted Likelihood Estimate (WLE) person-separation reliability was 0.53 and an Expected A Posteriori estimate based on Plausible Values (EAP/PV) reliability was 0.56, which was slightly lower than the acceptable value of 0.6. This suggested that some of the items in the original test might need to be reconsidered for inclusion in the test.

Table 1 summarizes the item properties of the IRT Rasch model and of the classical discriminant analysis. The items are listed from Q2 (i.e., item 2) to Q4 (i.e., item 4) according to their item difficulties ranging from 2.42 (most difficult) to -1.64 (least difficult) as well as their correct response rates ranging from 10% (lowest) to 81.33% (highest). The average person's proficiency was 0.00004 logits (*SD* = 0.93). The fit for single items (weighted mean squares, MNSQ) ranged from 0.89 to 1.12 (Mean = 1.00, *SD* = 0.06), thus indicating a good fit to the Rasch model at the item level. Finally, we applied point biserial correlations for the correct answers to obtain the classical discrimination values that ranged from 0.03 to 0.60.

In order to improve the reliability of the original version of the CTT-JH, each item was carefully examined based on the data reported in Table 1. First, Q10 was deleted due to its extremely low discrimination (0.03). Then, each of the remaining items was checked to ascertain whether the overall reliability would be increased when it was deleted. Finally, the best acceptable reliability of the overall test (EAP/PV reliability = 0.61) was obtained when Q12 and Q3 were deleted. Therefore, after deleting the three items Q10, Q12, and Q3, the final version of CTT-JH was formed with 12 items, as shown in Figure 3 and Table 2.

Table 1. Item properties of the Rasch model analysis and classical discrimination on the Junior High school computational thinking test (original 15 candidate items)

Item	Percent correct (%)	Difficulty	Discrimination	Infit MNSQ
Q2	10.00	2.42	0.35	0.97
Q12	14.44	1.98	0.23	1.03
Q10	16.00	1.81	0.03	1.12
Q6	16.89	1.76	0.39	0.97
Q14	29.11	1.00	0.28	1.05
Q3	39.78	0.47	0.30	1.06
Q9	43.56	0.29	0.40	1.00
Q1	44.67	0.24	0.37	1.02
Q15	47.78	0.10	0.39	1.02
Q13	55.11	-0.23	0.42	1.00
Q5	66.67	-0.78	0.60	0.89
Q11	72.44	-1.09	0.54	0.92
Q8	73.11	-1.12	0.52	0.92
Q7	79.33	-1.51	0.33	1.00
Q4	81.33	-1.64	0.43	0.96

Note. EAP/PV reliability = 0.56, WLE reliability = 0.53.

Figure 3. Wright Map of Rasch analysis on CTT-JH (12 items)

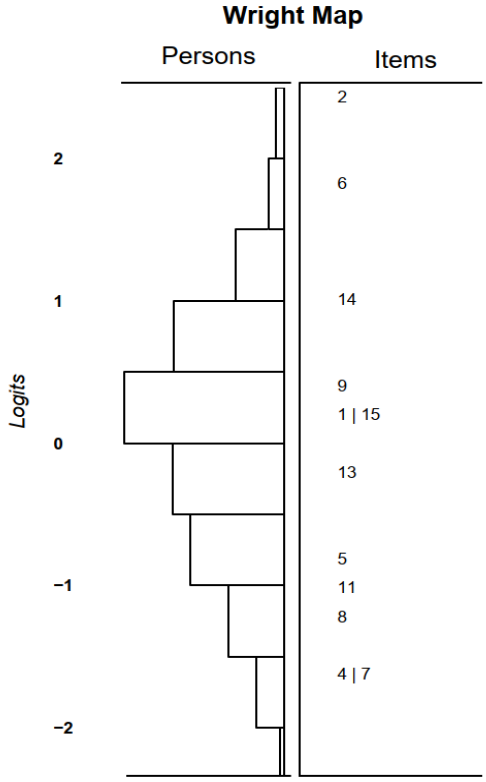


Figure 3 displays the distribution of students' abilities (on the left side) and item difficulties (on the right side) on the same logit scale. Items are ordered from the most difficult (at the top) to the least difficult (at the bottom). The histograms of students' abilities show that each student solved the item with a probability of 50% and was plotted from most able (at the top) to least able (at the bottom).

After we fitted the Rasch model, the results showed that the Expected A Posteriori estimate based on Plausible Values (EAP/PV) reliability was 0.61, Weighted Likelihood Estimate (WLE) person-separation reliability was 0.57, and Cronbach's alpha reliability was .6. Item difficulties ranged from 2.47 to -1.68. The average person's proficiency is 0.00018 logits (*SD* = 1.12). The fit for single items (weighted mean squares, MNSQ) ranged from 0.88 to 1.08 (Mean = 1.00, *SD* = 0.06), thus indicating a good fit to the Rasch model at the item level. In addition, we applied point biserial correlations for the correct answers to obtain the classical discrimination values that ranged from 0.28 to 0.60. The final list of items and its corresponding CT dimensions are shown in Table 3. The assessment items and the answering keys are available online at https://bit.ly/2022-CTT-JH.

Table 2. Item properties of Rasch analysis and classical discrimination on the final version of the CTT-JH (final 12 items)

Item	Percent correct (%)	Difficulty	Discrimination	Infit MNSQ
Q2	10.00	2.47	0.35	0.99
Q6	16.89	1.82	0.39	1.00
Q14	29.11	1.03	0.28	1.08
Q9	43.56	0.30	0.40	1.05
Q1	44.67	0.25	0.37	1.06
Q15	47.78	0.10	0.39	1.04
Q13	55.11	-0.24	0.42	1.01
Q5	66.67	-0.80	0.60	0.88
Q11	72.44	-1.11	0.54	0.91
Q8	73.11	-1.15	0.52	0.92
Q7	79.33	-1.54	0.33	1.03
Q4	81.33	-1.68	0.43	0.98

Note. EAP/PV reliability = 0.61, WLE reliability = 0.57, Cronbach's alpha = 0.6.

Table 3. Items of the CTT-JH responding to Selby and Woollard's (2013) CT framework

Item	Decomposition	Abstraction	Algorithm	Evaluation	Generalization
Q1. A toy goose is going out of farm			V		
Q2. The way to sell goose eggs			V		
Q4. Hungry goose open the boxes		V		V	
Q5. Transforming goose		V		V	
Q6. Let's shake hands after the match	V				
Q7. Best place for a bus stop				V	
Q8. Navigation app		V		V	V
Q9. Jack's code, QJ-Code	V				V
Q11. Swap the order and tell the secret			V		
Q13. The vine's weekly growth		V			V
Q14. Jack's self-driving car			V		V
Q15. Jack's henhouse management				V	V
Total items per dimension	2	5	4	4	4

3.2. The potential impact of gender and school resources on students' CT scores

Two-Way ANOVA was conducted to evaluate the effects of gender and different school ICT resources on the CT test mean scores. Homogeneity of variance of the four groups was verified according to Howell's study (2013, p. 234), which indicated that the results of variance analysis were more likely to be valid when the ratio of largest variance to smallest variance was four or below among the groups. In the current study, the ratio was 2.05 and revealed that the homogeneity assumption was not violated. Table 4 summarizes the two-way ANOVA results. No main effect for gender was observed (F = 1.71, p > .05, Partial eta squared < 0.01). However, school ICT resources reached significance on the CT test mean scores with medium to large effect size (F = 41.76, p < .01, Partial eta squared = 0.09). Most importantly, significant interaction with small to medium effect size occurred between gender and school ICT resources on the CT test mean scores (F = 5.86, p < .05, Partial eta squared = 0.01).

The regression line of students' school with more or fewer resources on CT mean scores was plotted for the different gender groups to better explain the interaction effects of Gender*School resources on CT test mean scores, as shown in Figure 4. To follow up on the significant interaction and to examine the differences between male and female students at different school resource levels, descriptive statistics and two independent *t* tests were used. Table 5 displays the results.

Table 4. Two-way ANOVA of the CT test scores

Source	df	MS	*F*	Partial eta squared
Gender	1	0.06	1.71	< 0.01
Schools with more or fewer ICT resources	1	1.35	41.76**	0.09
Gender * School ICT resources	1	0.19	5.86*	0.01

Note. df = degree of freedom; MS = Mean squares; $^{**}p < .01$; $^{*}p < .05$.

Firstly, the plots revealed that students' school level had a positive association with their CT test mean scores; that is, students in better resourced schools had higher CT test mean scores regardless of gender. Second, at schools with limited resources, male students had significantly lower CT test mean scores than female students did (male students' M = 0.41, SD = 0.20; female students' M = 0.48, SD = 0.14; t value = -2.37, $p < .05$). However, there was no significant difference in CT test mean scores between male and female students in schools with more resources (male students' mean scores = 0.57, standard deviation = 0.19; female students' mean scores = 0.55, standard deviation = 0.17; t value = 0.97, $p > .05$). Thirdly, the different slopes revealed that school ICT resources had a greater effect on students' CT test mean scores for male students than for female students. To summarize, it can be stated that school ICT resources, as well as the relationship of students' school ICT resources and gender, may be critical variables for CT test mean scores.

Figure 4. Interaction effect of gender and schools with more or fewer resources

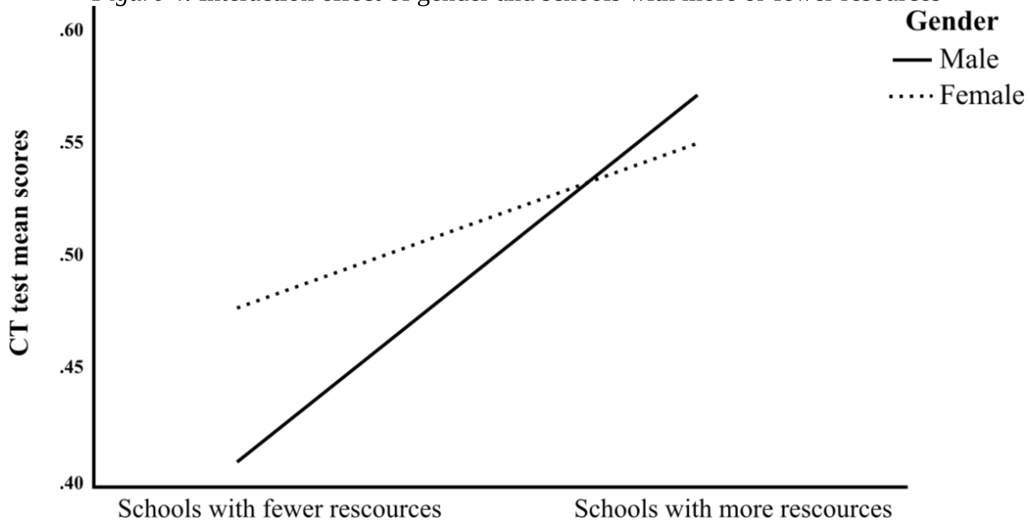


Table 5. Descriptive statistics for the independent variables

School levels	Gender	Mean	*SD*	*N*	*t*-test
Fewer resources	Male	0.41	0.20	77	-2.37*
	Female	0.48	0.14	65	
More resources	Male	0.57	0.19	157	0.97
	Female	0.55	0.17	138	

Note. $^{*}p < .05$.

4. Conclusions and discussion

The present study aimed to develop and validate a computational thinking test for junior high students. The results of IRT analysis showed that the revised version of the CTT-JH test is a reliable instrument for measuring junior high school students' computational thinking. While the Bebras Challenge tasks have been used worldwide, researchers have pointed out problems with the quality of the items (Hubwieser & Muhling, 2015). Nevertheless, only a few studies have provided robust evidence of the psychometric properties of this type of domain-general CT assessment instrument. Results of the current study show that through iteratively using the IRT Rasch model and the classical discriminant analysis, we were able to identify and remove unsuitable items. Our revised Bebras Challenge items, with attention to the wording, the representation, and the new context, represent a joint construct of domain-general CT. The final version including 12 items is suitable for measuring the CT competencies of students at junior-high school level.

The CTT-JH research instrument has potential applications and research implications in future studies. First, it can be used in both pretests and posttests as almost no prior instruction in particular discipline or logic training required. This paper-based instrument can also be adapted to assess students' learning gains in computer-free CT learning activities such unplugged computing robots. Second, because of the domain general nature this instrument, it is possible to assess students' CT across different disciplines and even in a transdisciplinary learning context. Some researchers have conceptualized itself CT as a transdisciplinary concept and the needs of integrative thinking skills (Li et al., 2020). While no commonly accepted definitions of integrative thinking are available and instruments for assessing integrative thinking are scarce (National Research Council, 2014), we argued that domain-general CT assessment can be used for measuring integrative thinking when it is used in transdisciplinary context such as STEM education.

The results also show that school ICT resources, as well as the interaction between school ICT resources and gender, may be critical variables for students' domain-general CT competencies. On the one hand, students in schools with more ICT resources had higher CT test mean scores regardless of gender. This finding not only supports prior research which found no gender differences in CT test results (Durak & Saritepeci, 2018), but more importantly, it implies that school ICT resources play some role in students' development of CT competencies even for the non-programming, domain-general CT. One possible explanation is that digital learning nowadays has been applied to different aspects of learning. When students study in an advanced ICT environment, they might have more access to ICT use for various cognitive tasks such as analyzing, creating, exchanging, and using data and information in different subject areas (Herselman & Britton, 2002). In another study, Sirakaya (2020) found that students' CT skills were associated with their internet experience, mobile device experience, and mobile internet experience. These findings support the association between CT competencies and ICT usage. Directly or indirectly, access to and use of ICT resources may have helped to develop students' domain-general CT and to close the digital divide (Rallet & Rochelandet, 2007). An important implication of this finding is that school ICT resources do not only impact students' computer literacy or programming learning, but may also influence students' domain-general CT competencies as one key competency in the 21st century. This is an important area that should be considered in future educational policy for school ICT funding.

On the other hand, at schools with limited resources, male students had significantly lower CT test mean scores than female students did. In other words, male more than female students' CT competencies are affected by the lack of ICT resources in schools. Attention to the interactions between gender and ICT provides another angle for possible explanations of why empirical evidence of gender differences was inconclusive. We hypothesized that there might be different models of how students develop CT in ICT-deprived versus ICT-advanced environments. In ICT-deprived learning environments, students' domain-general CT competencies might have strong relationships to academic achievement. Previous research reported that other academic skills such as mathematical thinking and reading and verbal skills (Zhang & Nouri, 2019; Roman-Gonzalez et al., 2018) were found to be related to students' CT competencies. Furthermore, in a previous meta-analysis study (Lei et al., 2020), researchers concluded that students' CT is correlated to school achievement; furthermore, the correlations are stronger among female than male students. In other words, in ICT-deprived schools, female students' better CT competencies than male students may be related to female students' overall school achievement.

Another possible explanation regards the gender differences of ICT usage outside of schools. Kim et al. (2021) surveyed 23,000 elementary and middle school students in Korea and found that female students had higher ICT literacy levels than male students. They attributed the ICT literacy difference to the different ICT usage habits and attitudes of males and females. For instance, researchers have found gender differences in Internet using purposes and intensity at high school level (Tsai & Tsai, 2010); female students tended to use the Internet for communication purposes while male students tended to use the Internet for exploration purposes. Moreover, female students are more likely to use ICT after school for learning or doing homework than male students (Ahn & Chae, 2016), and female students used ICT to gain more experience of problem solving through social networks while male students used ICT for entertainment and games (Sung & Choi, 2016). The aforementioned studies provide possible explanations as to why in the current study we found that female students outperformed male students. When schools have fewer ICT resources, students' habits of ICT usage outside school can become even more influential. How to help male students to gain ICT competencies in ICT-deprived learning environments and how to gain understanding of what causes the gender differences are important questions to be studied in the future.

Finally, we identified some limitations in this study. In the current study, we explored the impact of gender and school resource interactions but did not have data to identify the epistemic resources of students' CT competencies. Careful and in-depth inquiries into the gender differences in conjunction with school ICT and comprehensive data collection, such as including data of ICT usage in-class and outside of school are suggested for future research. Moreover, we were able to categorize the students' schools into fewer or more ICT resources by reviewing ICT-related information from the schools. Future studies can further develop a system to quantify the school-level ICT information or quantify student-level ICT usage and include the data in a more complex statistical model by using statistics such as hierarchical linear modeling (HLM). Perhaps it is not possible to fully understand the relationships among domain-general ICT, gender, and ICT resources without expanding the understanding to students' other academic competencies or students' ICT usage in daily life. Further investigations of this area are required in future research.

Acknowledgment

This study was supported by the Ministry of Science and Technology (MOST) of Taiwan under the following project numbers: MOST 106-2511-S-003-065-MY3, 109-2511-H-003-019-MY3, 109-2511-H-003-052-MY3, 109-2628-H-020-001-MY3 and 111-2410-H-003-130-MY3. It was also supported by the "Institute for Research Excellence in Learning Sciences" of NTNU sponsored by the Ministry of Education (MOE) in Taiwan.

References

Ahn, S. & Chae, K. (2016). Correlation analysis on ICT literacy level and difference of habit to use ICT. *Journal of the Korean Association of Information Education, 20*(3), 303-312. https://doi.org/10.14352/ jkaie.2016.20.3.303

Andrich, D., & Marais, I. (2019). *A Course in Rasch measurement theory*. Springer.

Angeli, C., & Giannakos, M. (2020). Computational thinking education: Issues and challenges. *Computers in Human Behavior, 105,* 106185. https://doi.org/10.1016/j.chb.2019.106185

Angeli, C., & Valanides, N. (2020). Developing young children's computational thinking with educational robotics: An Interaction effect between gender and scaffolding strategy. *Computers in Human Behavior, 105,* 105954. https://doi.org/10.1016/j.chb.2019.03.018

Araujo, A. L. S. O., Andrade, W. L., Guerrero, D. D. S., & Melo, M. R. A. (2019). How many abilities can we measure in computational thinking? A Study on Bebras challenge. In *Proceedings of the 50th ACM Technical Symposium on Computer Science Education* (pp. 545–555). https://doi.org/10.1145/3287324.3287405

Atmatzidou, S., & Demetriadis, S. (2016). Advancing students' computational thinking skills through educational robotics: A Study on age and gender relevant differences. *Robotics and Autonomous Systems, 75,* 661-670. https://doi.org/10.1016/j.robot.2015.10.008

Aydin, M. (2021). Does the digital divide matter? Factors and conditions that promote ICT literacy. *Telematics and Informatics, 58,* 101536. https://doi.org/10.1016/j.tele.2020.101536

Basu, S., Rutstein, D. W., Xu, Y., Wang, H., & Shear, L. (2021). A Principled approach to designing computational thinking concepts and practices assessments for upper elementary grades. *Computer Science Education, 31*(2), 169-198. https://doi.org/10.1080/08993408.2020.1866939

Brennan, K., & Resnick, M. (2012). New frameworks for studying and assessing the development of computational thinking. In *Proceedings of the 2012 annual meeting of the American educational research association* (Vol. 1, pp. 25).

Chiazzese, G., Arrigo, M., Chifari, A., Lonati, V., & Tosto, C. (2019). Educational robotics in primary school: Measuring the development of computational thinking skills with the Bebras tasks. *Informatics, 6*(4), 43. https://doi.org/10.3390/informatics6040043

Clarke-Midura, J., Silvis, D., Shumway, J. F., Lee, V. R., & Kozlowski, J. S. (2021). Developing a kindergarten computational thinking assessment using evidence-centered design: The Case of algorithmic thinking. *Computer Science Education, 31*(2), 117-140. https://doi.org/10.1080/08993408.2021.1877988

Çoban, E., & Korkmaz, Ö. (2021). An Alternative approach for measuring computational thinking: Performance-based platform. *Thinking Skills and Creativity, 42,* 100929. https://doi.org/10.1016/j.tsc.2021.100929

Cutumisu, M., Adams, C., & Lu, C. (2019). A Scoping review of empirical research on recent computational thinking assessments. *Journal of Science Education and Technology, 28*(6), 651-676. https://doi.org/10.1007/s10956-019-09799-3

Czerkawski, B. C., & Lyman, E. W. (2015). Exploring issues about computational thinking in higher education. *TechTrends, 59*(2), 57-65. https://doi.org/10.1007/s11528-015-0840-3

Dagienė, V., & Stupuriene, G. (2016). Bebras - A Sustainable community building model for the concept based learning of informatics and computational thinking. *Informatics in Education, 15*(1), 25-44. https://doi.org/10.15388/infedu.2016.02

Durak, H. Y., & Saritepeci, M. (2018). Analysis of the relation between computational thinking skills and various variables with the structural equation model. *Computers & Education, 116,* 191-202. https://doi.org/10.1016/j.compedu.2017.09.004

Eloy, A., Achutti, C. F., Fernandez, C., & Lopes, R. D. (2022). A Data-driven approach to assess computational thinking concepts based on learners' artifacts. *Informatics in Education, 21*(1), 33-54. https://doi.org/10.15388/infedu.2022.02

Fields, D., Lui, D., Kafai, Y., Jayathirtha, G., Walker, J. & Shaw, M. (2021) Communicating about computational thinking: Understanding affordances of portfolios for assessing high school students' computational thinking and participation practices. *Computer Science Education, 31*(2), 224-258. https://doi.org/10.1080/08993408.2020.1866933

Grover, S., & Pea, R. (2018). Computational thinking: A Competency whose time has come. In *Computer science education: Perspectives on teaching and learning in school* (pp. 19-38). Bloomsbury Academic.

Guenaga, M., Eguíluz, A., Garaizar, P. & Gibaja, J. (2021) How do students develop computational thinking? Assessing early programmers in a maze-based online game. *Computer Science Education, 31*(2), 259-289. https://doi.org/10.1080/08993408.2021.1903248

Guggemos, J. (2021). On the predictors of computational thinking and its growth at the high-school level. *Computers & Education, 161,* 104060. https://doi.org/10.1016/j.compedu.2020.104060

Herselman, M., & Britton, K. G. (2002). Analysing the role of ICT in bridging the digital divide amongst learners. *South African Journal of Education*, *22*(4), 270-274.

Hohlfeld, T. N., Ritzhaupt, A. D., Dawson, K., & Wilson, M. L. (2017). An Examination of seven years of technology integration in Florida schools: Through the lens of the levels of digital divide in schools. *Computers & Education, 113,* 135-161. https://doi.org/10.1016/j.compedu.2017.05.017

Howell, D. C. (2013). *Statistical methods for psychology*. Wadsworth Cengage Learning.

Hubwieser, P., & Mühling, A. (2015). Investigating the psychometric structure of Bebras contest: Towards measuring computational thinking skills. In P*roceedings of 2015 International Conference on Learning and Teaching in Computing and Engineering*. https://doi.org/10.1109/LaTiCE.2015.19

Irribarra, D. T., & Freund, R. (2014). *Wright Map: IRT item-person map with ConQuest integration.* http://github.com/david-ti/wrightmap.

International Society for Technology in Education (ISTE) & Computer Science Teachers Association (CSTA). (2011). *Operational definition of computational thinking for K-12 education*. https://cdn.iste.org/www-root/Computational_Thinking_Operational_Definition_ISTE.pdf

Izu, C., Mirolo, C., Settle, A., Mannila, L., & Stupurienė, G. (2017). Exploring Bebras tasks content and performance: A Multinational study. *Informatics in Education, 16*(1), 39-59. http://dx.doi.org/10.15388/infedu.2017.03

Kale, U., Akcaoglu, M., Cullen, T., & Goh, D. (2018). Contextual factors influencing access to teaching computational thinking. *Computers in the Schools, 35*(2), 69-87. https://doi.org/10.1080/07380569.2018.1462630

Kim, H. S., Kim, S., Na, W., & Lee, W. J. (2021). Extending computational thinking into information and communication technology literacy measurement: Gender and grade issues. *ACM Transactions on Computing Education (TOCE)*, *21*(1), 1-25. https://doi.org/10.1145/3427596

Kwon, K., Cheon, J., & Moon, H. (2021). Levels of problem-solving competency identified through Bebras Computing Challenge. *Education and Information Technologies, 26*(5), 5477-5498. https://doi.org/10.1007/s10639-021-10553-9

Lei, H., Chiu, M. M., Li, F., Wang, X., & Geng, Y. J. (2020). Computational thinking and academic achievement: A Meta-analysis among students. *Children and Youth Services Review, 118*, 105439. https://doi.org/10.1016/j.childyouth.2020.105439

Li, Y., Schoenfeld, A. H., diSessa, A. A., Graesser, A. C., Benson, L. C., English, L. D., & Duschl, R. A. (2020). On computational thinking and STEM education. *Journal for STEM Education Research, 3*(2), 147-166. https://doi.org/10.1007/s41979-020-00044-w

Light, J. (2009). Rethinking the digital divide. *Harvard Educational Review, 71*(4), 709-734. https://doi.org/10.17763/haer.71.4.342x36742j2w4q82

Mayer, D., Sodian, B., Koerber, S., & Schwippert, K. (2014). Scientific reasoning in elementary school children: Assessment and relations with cognitive abilities. *Learning and Instruction, 29*, 43–55.

Metcalf, S. J., Reilly, J. M., Jeon, S., Wang, A., Pyers, A., Brennan, K., & Dede, C. (2021). Assessing computational thinking through the lenses of functionality and computational fluency. *Computer Science Education, 31*(2), 199-223. https://doi.org/10.1080/08993408.2020.1866932

Moreno-León, J., Robles, G., & Román-González, M. (2015). Dr. Scratch: Automatic analysis of scratch projects to assess and foster computational thinking. RED. *Revista de Educación a Distancia,* (46), 1-23.

Mueller, J., Beckett, D., Hennessey, E., & Shodiev, H. (2017). Assessing computational thinking across the curriculum. In P. J. Rich & C. B. Hodges (Eds.), *Emerging research, practice, and policy on computational thinking, educational* *communications and technology: Issues and innovations* (pp. 251–267). Springer. https://doi.org/10.1007/978-3-319-52691-1_16

National Research Council. (2014). *STEM integration in K-12 education: Status, prospects, and an agenda for research.* National Academies Press. https://doi.org/10.17226/18612

Polat, E., Hopcan, S., Kucuk, S., & Sisman, B. A. (2021). Comprehensive assessment of secondary school students' computational thinking skills. *British Journal of Educational Technology*, 52, 1965-1980. https://doi.org/10.1111/bjet.13104

Poulakis, E., & Politis, P. (2021). Computational thinking assessment: Literature review. In T. Tsiatsos, S. Demetriadis, A. Mikropoulos, & V. Dagdilelis (Eds.), *Research on E-Learning and ICT in Education: Technological, Pedagogical and Instructional Perspectives* (pp. 111-128). Springer International Publishing. https://doi.org/10.1007/978-3-030-64363-8_7

Rallet, A., & Rochelandet, F. (2007). ICTs and inequalities: The Digital divide. In E. Brousseau, & N. Curien (Eds.), *Internet and digital economics: principles, methods and applications* (pp. 693–717). Cambridge University Press. https://doi.org/10.1017/cbo9780511493201.025

Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests.* Institute of Educational Research.

Robitzsch, A., Kiefer, T., & Wu, M. (2020). *TAM: Test analysis modules. R package version 2.6-2*. http://CRAN.R-project.org/package=TAM

Román-González, M., Moreno-León, J., & Robles, G. (2019). Combining assessment tools for a comprehensive evaluation of computational thinking interventions. In S.-C. Kong & H. Abelson (Eds.), *Computational Thinking Education* (pp. 79-98). Springer Singapore.

Román-González, M., Pérez-González, J.-C., & Jiménez-Fernández, C. (2017). Which cognitive abilities underlie computational thinking? Criterion validity of the Computational Thinking Test. *Computers in Human Behavior, 72,* 678-691. https://doi.org/10.1016/j.chb.2016.08.047

Román-González, M., Pérez-González, J.-C., Moreno-León, J., & Robles, G. (2018). Extending the nomological network of computational thinking with non-cognitive factors. *Computers in Human Behavior, 80,* 441-459. https://doi.org/10.1016/j.chb.2017.09.030

Selby, C., & Woollard, J. (2013). *Computational thinking: The Developing definition*. https://eprints.soton.ac.uk/356481/

Shute, V. J., Sun, C., & Asbell-Clarke, J. (2017). Demystifying computational thinking. *Educational Research Review, 22,* 142-158. https://doi.org/10.1016/j.edurev.2017.09.003

Sirakaya, D. A. (2020). Investigation of computational thinking in the context of ICT and mobile technologies. *International Journal of Computer Science Education in Schools, 3*(4), 50-59. https://doi.org/10.21585/ijcses.v3i4.73

Sung, E. & Choi, H. (2016). Exploring the types of classes and characteristics on ICT literacy of middle school students with latent class analysis. *Journal of Educational Technology, 32*(4), 987-1013. http://doi. org/ 10.17232/KSET.32.4.987

Tang, X., Yin, Y., Lin, Q., Hadad, R., & Zhai, X. (2020). Assessing computational thinking: A Systematic review of empirical studies. *Computers & Education, 148,* 103798. https://doi.org/10.1016/j.compedu.2019.103798

Tsai, M.-J., & Tsai, C.-C. (2010). Junior high school students' Internet usage and self-efficacy: A Re-examination of the gender gap. *Computers & Education, 54*(4), 1182-1192. https://doi.org/10.1016/j.compedu.2009.11.004

Tsai, M.-J., Chien, F. P., Lee, S. W.-Y., Hsu, C.-Y., & Liang, J.-C. (2022). Development and validation of the computational thinking test for elementary school students (CTT-ES): Correlate CT competency with CT disposition. *Journal of Educational Computing Research, 60*(5), 1110-1129. https://doi.org/10.1177/07356331211051043

Tsai, M.-J., Liang, J.-C., & Hsu, C.-Y. (2021). The Computational thinking scale for computer literacy education. *Journal of Educational Computing Research, 59*(4), 579-602. https://doi.org/10.1177/0735633120972356

Tsarava, K., Moeller, K., Román-González, M., Golle, J., Leifheit, L., Butz, M. V., & Ninaus, M. (2022). A Cognitive definition of computational thinking in primary education. *Computers & Education, 179,* 104425. https://doi.org/10.1016/j.compedu.2021.104425

van der Vegt, W. (2018). How hard will this task be? Developments in analyzing and predicting question difficulty in the Bebras Challenge. *Olympiads in Informatics, 12,* 119-132. http://dx.doi.org/10.15388/ioi.2018.10

Voogt, J., Fisser, P., Good, J., Mishra, P., & Yadav, A. (2015). Computational thinking in compulsory education: Towards an agenda for research and practice. *Education and Information Technologies, 20*(4), 715-728. https://doi.org/10.1007/s10639-015-9412-6

Weintrop, D., Rutstein, D. W., Bienkowski, M., & Mcgee, S. (2021). Assessing computational thinking: An Overview of the field. *Computer Science Education, 31*(2), 113-116. https://doi.org/10.1080/08993408.2021.1918380

Wing, J. M. (2006). Computational thinking. *Communications of the ACM*, *49*(3), 33-35.

Zhang, L. & Nouri, J. (2019). A Systematic review of learning computational thinking through Scratch in K-9. *Computers & Education*, *141*, 103607. https://doi.org/10.1016/j.compedu.2019.103607

Zhang, M. (2014). Who are interested in online science simulations? Tracking a trend of digital divide in Internet use. *Computers & Education*, *76,* 205-214. https://doi.org/10.1016/j.compedu.2014.04.001

Appendix

ICT resource information for each school where data were collected

School[a]	Location	(a) Number of full-time ICT teacher in school	(b) Number of classes in school	(c) = (a) / (b) Full-time ICT teacher-class ratio	(1) ICT teacher-class ratio (if c > 0.1 coded 1, otherwise coded 0)	(2) Funding of Maker Education and Technology Center (Yes = 1, Not Available = 0)	(3) ICT-integrated curriculum project (Yes=1, Not Available=0)	Total Score = (1)+(2)+(3)	ICT resources in school (More: Total Score > 0, Fewer: Total Score = 0)
School A	City	3	61	0.049	0	1	0	1	More resource
School B	City	2	33	0.061	0	0	0	0	Fewer resource
School C	County	0	24	0.000	0	0	0	0	Fewer resource
School D	Rural	3	27	0.111	1	0	1	2	More resource
School E	Rural	6	43	0.140	1	1	0	2	More resource
School F	Remote	0	28	0.000	0	1	1	2	More resource

Note. [a]Data were obtained from the Ministry of Education of Taiwan during 2019

Lin, Y.-T. (2023). Learning Performances towards the BookRoll E-Book System for Flipped Classrooms in Software Engineering Education. *Educational Technology & Society*, *26*(3), 190-202. https://doi.org/10.30191/ETS.202307_26(3).0014

Learning Performances towards the BookRoll E-Book System for Flipped Classrooms in Software Engineering Education

Yen-Ting Lin
Department of Computer Science and Artificial Intelligence, National Pingtung University, Taiwan // ricky014@gmail.com

(Submitted June 26, 2022; Revised December 7, 2022; Accepted January 28, 2023)

ABSTRACT: The aim of software engineering education is to educate students in software technologies, developments, procedures, and scientific practices to enable them to cope with industrial demands. However, the implementation of software engineering education in traditional university classrooms is restricted by the semester structure, making it difficult to achieve a proper learning balance between theory and practice. To balance theoretical and practical learning, prior studies have indicated that flipped learning is a suitable classroom setting for students and teachers. In a flipped learning environment, it is important to enhance and capture students' learning performance before the class to facilitate teachers and students in proceeding with in-class instruction and learning. In this study, an e-book system named BookRoll was applied to support software engineering education in a flipped learning setting. The proposed approach supports and facilitates out-of-class and in-class learning by providing reading and learning analytic functions for teachers and students. To evaluate the proposed approach, two classes of students were allocated to an experimental group and a control group to participate in an experiment. In the flipped learning process, the experimental group was supported by the BookRoll system, while the control group did not use the BookRoll system. The results revealed that the proposed approach not only promoted students' learning achievements in software engineering education but also improved their learning motivation, attitude, and problem-solving ability. The reading behavior analysis further indicated that reading time was a statistically significant predictor of learning achievement.

Keywords: Flipped classroom, Software engineering education, E-book system, Reading behaviors, Quality education

1. Introduction

Software is an abstract object that is quite different from physical and digital artifacts. Before software is compiled and successfully deployed, nobody can view, touch, or experience it. Therefore, during software development processes, engineers may encounter unexpected or problems or uncertainty, increasing development risks and costs. To reduce such risks and costs, it is important to apply scientific approaches to design and develop software effectively and efficiently. The major aim of software engineering is to encourage software engineers to apply scientific methodologies to efficiently and effectively develop high-quality software (Cico et al., 2021). Therefore, software engineering education is necessary to teach students the relevant knowledge and skills required for the whole software life cycle. Software engineering education has evolved over the past 30 years, but addressing proper learning with theory and practice is still an open issue (Lin, 2021). The reason is that the implementation of software engineering education in traditional university classrooms is restricted by semester settings, making it difficult to balance theoretical and practical learning activities in traditional university classes (Baker et al., 2005). This means there is scant opportunity to enable students to thoroughly learn software engineering (Garousi et al., 2020).

To balance theoretical and practical learning, prior studies have indicated that flipped learning is a suitable classroom setting for teachers and students (Lee et al., 2021). A flipped classroom is a student-centered pedagogical approach that flips traditional in-class instruction and out-of-class homework. In traditional lecture-based classrooms, students usually carry out lower-level cognitive learning in class, while they are required to perform higher-level cognitive work outside class. In flipped classroom settings, students are usually asked to participate in online learning to acquire knowledge before the class and then to further engage in practical and interactive activities to learn higher-order thinking skills in class. The literature has indicated that in flipped learning, students' learning performance outside class is positively related to their learning performance in class (Birgili et al., 2021; van Alten et al., 2020). Moreover, enhancing and capturing students' learning performance before the class is important to facilitate teachers and students in proceeding with in-class instructions. However, the traditional flipped learning approach cannot support teachers and students conducting learning activities outside the class since students cannot obtain effective learning tools and teachers cannot effectively capture students' learning status. Thus, teachers and students may conduct in-class learning activities on faulty foundations, which may further negatively affect students' learning performance.

In this study, an e-book system named BookRoll was applied to support software engineering education in a flipped classroom setting. The proposed approach supports and facilitates out-of-class and in-class learning by providing reading and learning analytic functions for students and teachers. To explore the performance of the proposed approach, an experiment was designed to investigate students' learning motivation, learning attitude, learning achievement, and problem-solving ability in software engineering education. In addition, students' reading behaviors in the proposed approach were analyzed and discussed.

2. Literature review

2.1. Flipped classroom

Over the past decade, the flipped classroom has been applied at different education levels and in different subjects (Esperanza et al., 2021; Gilboy et al., 2015; Zou, 2020). The flipped classroom was first proposed in 2007 (Roehl et al., 2013; Zengin, 2017). Moreover, numerous studies have presented and advocated the positive effects of flipped learning approaches on students' learning performance (Abdullah et al., 2019; Martínez-Jiménez & Ruiz-Jiménez, 2020; Hwang & Chang, 2020). Furthermore, the literature has reported that students expressed a preference for flipped classrooms over traditional classrooms (Lew, 2016; McNally et al., 2017). Several studies have also indicated that students' engagement is important for improving teaching and learning effectiveness in flipped classrooms (Barkley, 2010; Coates, 2006). On the other hand, investigations have claimed that inappropriate flipped learning settings have a negative effect on students' engagement during the learning process (Patanwala et al., 2017; DeRuisseau, 2016). Related works have concluded that flipped classrooms benefited students' learning performance because the students had enough time to engage in higher-level cognitive learning in class (Lo et al., 2017; Chung et al., 2021). This finding is in accord with the results of Bryson and Hand (2007), who found that students were more likely to engage in high-level cognitive activities with teachers' support.

Previous studies have also pointed out that technology-enhanced flipped classrooms can improve students' learning achievement, facilitate their course participation and satisfaction, and enhance their confidence, creativity and problem-solving ability (Akçayır et al., 2018; Lin, 2016; Yang et al., 2021). Huang et al. (2022) asked students to watch videos and teaching materials before class to understand specific content to be discussed in class. The students in the control group learned with the flipped classroom learning strategy, while the students in the experimental group learned with the flipped learning strategy assisted by business simulation games. The results indicated that the technology-assisted flipped learning approach had a significant positive impact on students' engagement, higher-order thinking skills, and learning performance. Ye et al. (2019) proposed a flipped learning system and used an interactive problem-posing guiding approach to facilitate students' learning before class. The results showed that students who used the approach of the preclass preview had improved learning performance and self-efficacy. Huang et al. (2023) integrated an AI-enabled personalized video recommendation system into a systems programming course in a flipped classroom setting. In the proposed approach, learning videos were recommended according to each student's learning process so that students could learn before and after class. The results showed that the proposed approach improved students' learning engagement and performance with a moderate motivation level.

In addition to the benefits mentioned above, some research issues regarding flipped classrooms remain open. A major issue is the learning performance of students before class, as it affects how teachers and students conduct learning activities in the classroom. In flipped classroom learning environments, students may spend more time previewing before class than the time spent in class. Therefore, the results of previewing before class are very important for learning performance. Most previous studies have focused on the design of classroom learning activities and how to maintain students' participation in preclass learning and improve their self-learning ability (Bond, 2020; Lai et al., 2021; Rasheed et al., 2020). Another challenge is that teachers cannot effectively capture students' learning status before they attend the class. In this circumstance, students may not be able to follow the coursework presented to them in class, or teachers may not proceed with the course as planned. Students with faulty foundations could thus be at risk in taking on learning activities in class. Therefore, it is important to apply an appropriate learning system to support students' out-of-class learning and to assist teachers in capturing students' learning status in flipped classrooms (Hwang & Lai, 2017; Yang et al., 2021). There is a lack of research that allows teachers to effectively understand students' learning behaviors outside the classroom.

2.2. E-book

In the past two decades, with the development and increasing popularity of computer and mobile technology, a trend has emerged of developing e-book resources and technologies to aid students in more efficiently reading, studying, and interacting with learning content. An e-book presents a document in digital form with digital resources and services on certain content platforms, learning systems, or reading software using a computer or mobile device. E-books present the advantages of reading with information technology, such as digitalized bookmarks, annotations, notes, and queries. Moreover, e-books provide digitalization, interaction, and individualization functions to help students read learning materials (Chen & Su, 2019). Several investigations have indicated that these reading technologies can be helpful for students' reading and learning performance (Lin et al., 2013; Sung et al., 2022; Zhao et al., 2021). For instance, Connor et al. (2019) developed e-books to improve primary school students' reading and help them develop a better understanding of words. The results indicated that e-books could enhance students' learning performance and support their development of metacognition. Furthermore, Ni'mah and Umamah (2020) applied e-books to improve students' English skills. The results showed that e-books helped the students improve their English skills and develop reading habits. In addition, the study found that the students had positive perceptions while performing digital reading.

In addition to enhancing students' learning performance, related studies have indicated that a significant amount of logged reading data can be tracked by reading e-books, while those data are much more difficult to track when reading paper-based books (Merkle et al., 2022; Lim et al., 2021). The literature has also shown that learning behaviors and log data are positively related to students' efforts, performances, and outcomes (Huang et al., 2020; Al-Ahdal, 2020). Several studies have reported that the analysis of students' reading behaviors can assist teachers in designing learning materials or learning activities (Sutcliffe & Hart, 2016; Yin et al., 2019). Mouri et al. (2018) proposed an e-book system to support the visualization and analysis of students' reading logs to improve language learning. Boticki et al. (2019) showed that e-books can support teachers in monitoring students' learning status during course delivery and identifying at-risk students early.

Several studies have reported a positive correlation with students' learning performance in different courses by adopting flipped classrooms with e-books (Mukhlisa et al., 2021; Siswanto, 2021; Palinussa et al., 2021). These results have provided evidence that students who learn with e-books can achieve high learning performance in flipped classrooms. The literature has also emphasized that students in flipped classroom settings who used e-books exhibited significant engagement in preclass preparation (Fahmi et al., 2020; Dembedza & Chipurura, 2020). These findings supported flipped classrooms with e-books as a successful approach to improve students' learning performance out of class. Based on the literature review, to promote software engineering education, this study conducted software engineering courses in flipped learning settings. Moreover, an e-book system was adopted to facilitate students' learning in the flipped classroom.

2.3. The importance of learning motivation, learning attitude, and problem-solving ability to learning

Learning motivation refers to the psychological process of arousing students' learning activities, maintaining learning activities, and enabling students to move toward learning goals (Ng, 2018; Ryan & Deci, 2019). Motivation affects students' participation in classroom learning as well as their degree of hard work and perseverance in completing tasks. Students with higher learning motivation will show higher participation in courses and greater persistence in completing tasks. Learning motivation can drive students to study hard automatically and spontaneously and then improve their learning performance. It is a very important factor in the learning process (Wu & Wu, 2020). Therefore, improving learning motivation is very important for learning.

Learning attitude is composed of three components: cognitive, affective, and behavioral intentions, which are embodied in students' attitudes toward courses, teaching materials, teachers, and the school environment (Svenningsson et al., 2022). The cognitive component refers to students' knowledge and understanding of learning activities or courses, which reflects their inner evaluation of learning and is the basis of their learning attitude. The affective component is the emotions or emotional experiences that students produce through cognition. It is the core component that affects learning attitude. The behavioral intention component refers to students' tendency to respond to learning, which is affected by cognitive and affective components, and then engage in certain behaviors (Svenningsson et al., 2022).

Malik et al. (2022) indicated that in programming courses, in addition to focusing on the knowledge teaching of programming logic, students should develop their problem-solving ability. Because the development of problem-solving ability helps students understand the input, procedure, output and other requirements required by the problem task, it has a positive impact on learning behavior and learning attitude. Matturro et al. (2019)

mentioned that team members participating in software development projects become proficient not only in programming skills, such as methods, tools and technologies, but also in problem-solving ability. Shanta and Wells (2022) found that having good problem-solving ability can promote students' understanding of knowledge. Therefore, problem-solving ability is very important in improving students' learning performance.

3. Methodology

To explore the effects of the proposed approach on students' learning performance in software engineering education, this study employed a nonequivalent-groups quasi-experimental design at a Taiwanese university. The scope of software engineering education is related to the software life cycle, which covers system requirement analysis, system design and implementation, system testing and validation, and system maintenance and evolution.

3.1. Participants

The participants were third-year and fourth-year students (aged 20-22 years) from two software engineering classes in the department of computer science at the university. All participants were assigned to two groups according to their class. In all, 64 students were recruited. Twenty-seven students and 37 students were allocated to the control group and experimental group, respectively. Students in the control group participated in software engineering education supported by the flipped learning approach without the e-book system. Students in the experimental group engaged in software engineering education supported by the flipped learning approach with the e-book system. The students in both groups were taught by a teacher with more than 10 years of teaching experience in software engineering education.

3.2. BookRoll E-book system

The BookRoll system was developed by the Ogata Laboratory at Kyoto University (Ogata et al., 2017; Ogata et al., 2015). BookRoll is not only an e-book reading system but also a reading tracker. It has been applied in several academic studies to address different educational issues (Chen et al., 2019; Mouri et al., 2018; Yang et al., 2021). The system is a web application that allows teachers to manage digital learning resources (including e-books and reading status) and enables students to use a web browser to read and mark the e-books anytime and anywhere. Reading functions such as bookmarking, highlighting, e-notes, and searches are provided to support students' reading activities. During the reading process, students' reading behaviors while using the system, including notes, page movements, highlights, bookmarks, and reading progress, are captured and stored in the system database. The e-book system also provides a dashboard to enable teachers to easily and systematically capture aggregated information about students' reading engagement and activity. Moreover, the system integrates several analyzers and aggregators into the dashboard to display students' reading behavior information for teachers, including the number of notes, note contents, the number of red and yellow markers on each page, the degree of reading completion, the contents of highlighted markers, page transition status, and event rate and total numbers over a range of time.

3.3. Instruments

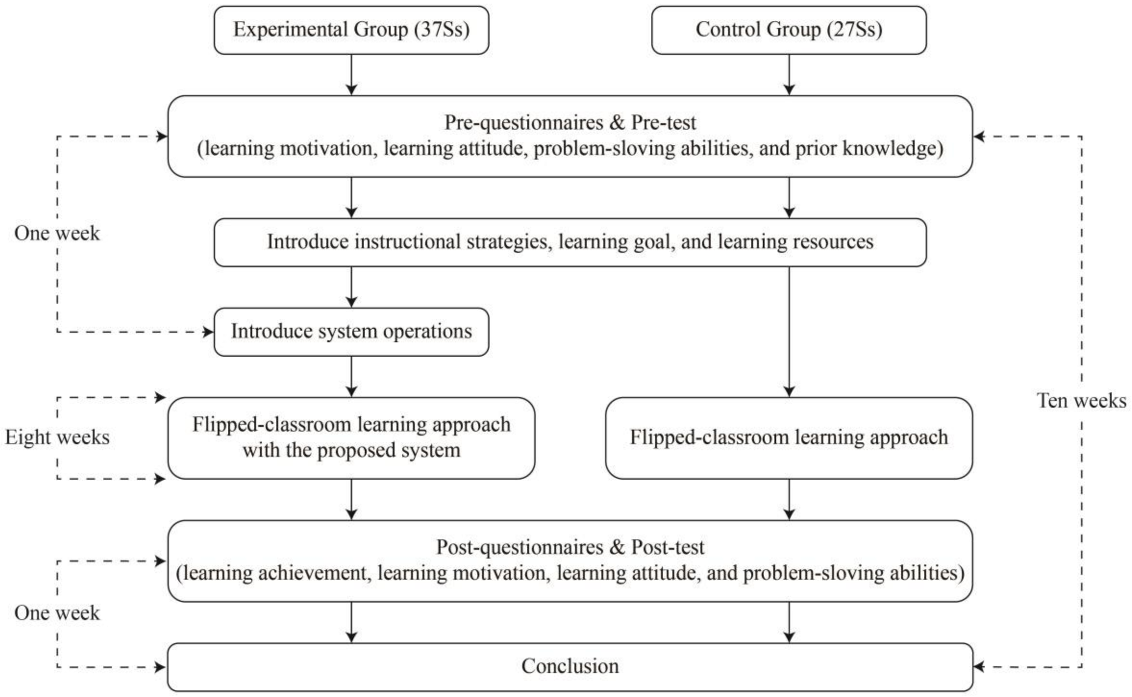
The research instruments used in this study were a learning motivation questionnaire, learning attitude questionnaire, problem-solving ability questionnaire, prior knowledge test, and learning achievement test. In addition, students' reading behaviors collected by the e-book system were analyzed. The prior knowledge test was used to measure students' prior knowledge of software engineering before taking the course. The learning achievement test was used to evaluate students' learning achievement after completing the course. The two tests were conducted as paper-and-pencil tests, and the highest possible score on the two tests was 100. The two tests have been used to evaluate students' knowledge level of software engineering in several software engineering classes (Lin, 2019; Lin, 2021). The reliability, difficulty, and discrimination of the tests have been verified.

With regard to measuring students' learning motivation, the scale of intrinsic value in the Motivated Strategies for Learning Questionnaire (MSLQ) was used as the learning motivation questionnaire with nine items in this study (Erturan Ilker et al., 2014; Lin & Cheng, 2022; Pintrich & De Groot, 1990). Each item was scored on a 7-point Likert scale from 1 (strongly disagree) to 7 (strongly agree). For all students participating in this experiment, the Cronbach's alpha values of the learning motivation questionnaire in the pretest and posttest were 0.919 and 0.948, respectively. The learning attitude questionnaire was proposed by Hwang and Chang (2011) and has been used to survey students' learning attitudes in several studies (Huang & Hwang, 2019; Lin et al., 2018; Liu et al., 2021). It has seven items and uses a 4-point Likert scale ranging from 1 (strongly disagree) to 4 (strongly agree). The reliability values of the pretest and posttest learning attitude questionnaires were 0.849 and 0.922, respectively. To survey students' problem-solving ability, 25 items were referenced and adopted in this study (Lin, 2019; Lin, 2021). Each item was scored on a 5-point Likert scale ranging from 1 (totally disagree) to 5 (totally agree). The Cronbach's alpha values of the problem-solving skills questionnaire in the pretest and posttest were 0.856 and 0.845, respectively.

3.4. Experimental procedure

The experiment of the two groups had a length of ten weeks (27 hours), including eight weeks (24 hours) of participating in the software engineering courses and two weeks of taking the pretests and posttests (3 hours). As shown in Figure 1, during the first week, the students in both groups completed three prequestionnaires to survey individual learning motivation learning attitude and problem-solving ability before participating in the formal instruction. Then, all students were asked to complete a prior knowledge test to evaluate their software engineering knowledge. The teacher then spent 30 minutes introducing the flipped learning approach to the students. The students in the experimental group received an additional 20-minute introduction to learn the operation of the e-book system.

Figure 1. The experimental process



During the following eight weeks, the teacher assigned the same e-books and videos to the students in both groups every week. Moreover, the teacher asked each student who had to read and watch the relevant e-books and videos to learn the weekly knowledge prior to attending the class. For this reason, 10 e-books and 61 videos (each approximately 5 to 8 minutes) were produced for this study to support the software engineering courses. The students in the control group used a PDF reader and YouTube player to read and watch the e-books and videos before the weekly class. In contrast, the students in the experimental group used the e-book system and YouTube player to read and watch the e-books and videos before the weekly class. The students in the experimental group were asked to use the bookmark, highlight, and e-note functions of the e-book system to read the weekly e-books. By using the system, the students' reading behaviors, including notes, page movements, highlights, bookmarks, and reading progress, were captured and stored in the database. The teacher could use the dashboard function of the e-book system to capture aggregated information about the students' reading status, such as highlights and notes on each page and reading completion. The teachers could then adjust in-class learning activities according to the students' reading status. In the classroom, the teacher arranged the students in the two groups to conduct discussions, case studies, and practices related to the scope of the software life cycle. During the last week of the experiment, the students in both groups received three postquestionnaires and a posttest to explore individual learning motivation, learning attitude, problem-solving ability, and learning achievement after completing all learning activities.

4. Results

4.1. Learning achievement analysis

This analysis was conducted to investigate the difference between the two groups in terms of learning achievement. To perform the analysis, the equivalent of the students' software engineering prior knowledge was first examined by an independent sample t-test. The mean value and standard deviation of the students in the control group were 34.00 and 17.32, respectively. Those values of the students in the experimental group were 36.21 and 14.59, respectively. To confirm the normal distribution of the data, the Kolmogorov-Smirnov test (KS test) was conducted because the number of samples in the analysis was larger than 50. The KS test result (F = .121, p > .05) indicated that the data were normal. Then, the homogeneity of variance was assessed by Levene's test, and the results showed that the within-group variances were equal (F = .868, p = .355 > .05). The results of the independent sample t-test showed no significant difference between the two groups with regard to prior knowledge of software engineering (t(1, 62) = .644, p = .589 > .05).

To check for the difference in learning achievement depending on the different flipped learning approaches among the two groups, a one-way ANCOVA was employed after the impact of the students' prior knowledge was neutralized. The assumption of homogeneity of the regression slope (F = 1.038, p > .05) was not violated, revealing that the ANCOVA was suitable. Table 1 tabulates the ANCOVA results for learning achievement for the two groups. The average learning achievement score was significantly higher for the experimental group than for the control group (F(1, 61) = 10.431, p = .002 < .05). The results show that in terms of learning achievement, the flipped learning approach with the e-book system was better for students than the flipped learning approach without the e-book system.

Table 1. ANCOVA results of learning achievement for the two groups

Group	n	Mean	Standard deviation	Adjusted mean	Adjusted Standard deviation	F	p-value
Experimental group	37	89.18	14.01	88.99	2.33	10.431	.002*
Control group	27	76.80	15.19	77.09	2.84		

Note. *p < .05.

4.2. Learning motivation analysis

In this analysis, one-way ANCOVA was used to determine whether students in the two flipped learning approaches had significant differences in the learning motivation scores on the postquestionnaire. ANCOVA was employed with the scores of the learning motivation prequestionnaire, the scores of the learning motivation postquestionnaire, and the learning approach as the covariate, dependent variable, and independent variable, respectively. The test of homogeneity showed that the assumption of regression homogeneity was not violated, with F = 1.127 and p = .293 > .05, indicating that ANCOVA can be used for learning motivation analysis.

Table 2 tabulates the ANCOVA results, which indicate that the post learning motivation scores of the two groups were significantly different, with F(1, 61) = 4.140 and p = .046 < .05. That is, the students in the experimental group had significantly higher learning motivation than those in the control group, indicating that the flipped learning approach with the e-book system was beneficial to the students' learning motivation in the software engineering course.

Table 2. ANCOVA results of learning motivation for the two groups

Group	n	Mean	Standard deviation	Adjusted mean	Adjusted Standard deviation	F	p-value
Experimental group	37	6.06	0.82	6.07	0.12	4.140	.046*
Control group	27	5.67	0.69	5.66	0.15		

Note. *p < .05.

4.3. Learning attitude analysis

This analysis was conducted to evaluate the effects of the proposed learning approach on students' learning attitudes toward software engineering education. ANCOVA was employed with the scores of the learning attitude prequestionnaire, the scores of the learning attitude postquestionnaire, and the learning approach as the covariate, dependent variable, and independent variable, respectively. The test of homogeneity showed that the assumption of regression homogeneity was not violated, with F = .353 and p = .555 > .05, revealing that ANCOVA can be used to analyze learning attitude scores.

Table 3 presents the ANCOVA results. The results show a significant difference between the two groups in the learning attitude postquestionnaire ($F(1, 61) = 7.524$, $p = .008 < .05$). This means that the use of the flipped learning approach with the e-book system positively affected the students' learning attitude. More specifically, students who engaged in the flipped learning approach with the e-book system significantly benefited in terms of learning attitude compared to those who engaged in the flipped learning approach without the e-book system.

Table 3. ANCOVA results of learning attitude for the two groups

Group	n	Mean	Standard deviation	Adjusted mean	Adjusted Standard deviation	F	p-value
Experimental group	37	3.54	.44	3.55	.07	7.524	.008*
Control group	27	3.25	.41	3.24	.08		

Note. $^*p < .05$.

4.4. Problem solving ability analysis

One-way ANCOVA was utilized to measure the students' problem-solving ability using the scores of the problem-solving ability prequestionnaire as covariate, the scores of the problem-solving ability postquestionnaire as a dependent variable, and the learning approach as an independent variable. The test of homogeneity showed that the assumption of regression homogeneity was not violated (F = .634, p > .05), indicating that ANCOVA can be used to conduct problem-solving ability analysis.

Examination of the effectiveness of the flipped learning approach with the e-book system in terms of problem-solving ability through ANCOVA showed that there was a significant difference between the experimental group and the control group, with F = 4.755 (p = .033 < .05), as shown in Table 4. The results indicate that the statistically significant effect (α = .05) on the problem-solving ability postquestionnaire was attributable to the learning approach, as the students in the experimental group, who learned in the flipped learning approach with the e-book system, performed better than their counterparts in the control group, who learned in the flipped learning approach only.

4.5. Reading behavior analysis

This study further analyzed the reading behaviors of students in the experimental group using reading logs from the e-book system. The reading logs covered reading time and reading completion. The reading time indicates that a student spends a certain number of minutes reading the e-books during the learning process. The reading completion is a student's average reading completion of all e-books during the entire learning process.

Table 4. ANCOVA results of problem-solving ability for the two groups

Group	n	Mean	Standard deviation	Adjusted mean	Adjusted Standard deviation	F	p-value
Experimental group	37	3.71	.43	3.72	.06	4.755	.033*
Control group	27	3.50	.21	3.49	.08		

Note. $^*p < .05$.

The simple linear regression analysis method was utilized to predict students' learning achievement based on their reading time and reading completion. Students' learning achievement is the dependent variable. Students' reading time and reading completion are independent variables. The regression coefficient values and analytic results are presented in Table 5. The results indicate that there was a significant positive correlation between reading time and learning achievement in this study. This means that as students' reading time increased, their learning achievement also increased. It is noteworthy that reading completion did not significantly predict student learning achievement.

Table 5. Analyzing correlations among reading behaviors and learning achievement by linear regression

Predictor	R^2	Adjusted R^2	B	t	p
Reading time	.578	.313	.104	4.002	$.000^*$
Reading completion	.281	.050	.171	1.657	.107

Note. $^*p < .05$.

5. Discussion and conclusion

The proposed flipped learning approach provides students with better learning performance in acquiring software engineering knowledge and better supports teachers in monitoring students' learning status in a flipped learning process. Furthermore, in terms of learning out of class, the students benefited more and were more motivated to conduct the learning activities. The contribution of the proposed approach is twofold. First, it helps students improve their software engineering learning performances through using the e-book system to efficiently construct relevant knowledge out of class and further engage in mastering the knowledge during the in-class case studies, discussions, and practices in the flipped learning setting. Second, this approach helps teachers improve flipped learning instruction by using the dashboard of the e-book system to effectively and efficiently monitor students' learning status before the class. In this study, to monitor the students' weekly learning status out of class, the teacher first observed the reading rate of the e-books of the whole class from the system dashboard before the students attended the class. Moreover, the teacher further observed the students' notes and highlights on each page of the e-books to capture what the students may not have understood. Furthermore, the average reading time for each page in the e-books could be used by the teacher to judge whether the students had encountered difficulties in learning the contents. By using various learning analytic functions provided by the e-book system, teachers can effectively and efficiently capture students' learning status before class and make adjustments to learning activities in class in real time.

The experimental results support the positive impacts of the proposed approach on students' learning motivation, learning attitude, learning achievement, and problem-solving ability in software engineering education. The results support the contention of Bergmann et al. that the application of appropriate technologies is a significant element to positively motivate students' learning and affect their attitude in flipped classrooms (Bergmann & Sams, 2012; Lin, 2019; Oweis, 2018). This study used a web application (the e-book system) that provides suitable reading functions to enable students to use a web browser to effectively and efficiently study course materials anytime and anywhere. The increased motivation and attitude can also be attributed to flipped learning programs that can be reinvented to satisfy the individual needs of students in a program (Lin, 2021). Horn (2011) argued that a student's interest is motivated when she or he can identify with the materials that are used in the learning process. Regarding students' learning achievement, the results of the proposed approach were significantly better than those for students learning without the e-book system in a flipped classroom. This discovery conforms to the study of Haghighi et al. (2019) where students who learned with appropriate tools before class could enhance their engagement in class and further improve their learning achievement in a flipped classroom. The argument also supports the results for students' problem-solving ability found in this study. The analysis showed that students learning with the proposed approach had higher problem-solving ability than those learning with only the flipped learning setting. This discovery supports the studies of Chang and Hwang (2018) and Kurnianto et al. (2019) in that when students build fundamental knowledge out of class in a flipped learning environment, they are able to engage in in-class learning activities and better develop higher-order thinking abilities, such as problem-solving ability. Students with good prior knowledge are able to better articulate concepts, illustrate and construct interrelationships among concepts, and generate higher levels of thinking (Alamri, 2019; Cai & Gu, 2022; Mamun et al., 2020).

Despite the valuable contributions of the research results, the present study also had some limitations that should be acknowledged. First, as a result of university semester considerations, this study did not adopt random selection to distribute the participants in the control and experimental groups. Second, the sample size was not large, so the results cannot be generalized to possible learning performances in different contexts and applications. In addition, this study was conducted in an Asian country. The research settings and results might not be directly applicable to software engineering classes in other countries with different cultures, educational policies, or learning environments. According to the literature review, most of the research related to flipped classrooms still focuses on the learning performance of students before class. However, future research should focus on what kind of instructional design framework to use to plan the entire flipped classroom approach rather than just emphasizing preclass learning (Lo & Hew, 2017). In other words, researchers can further design in-class learning activities to teach students higher-level thinking ability based on the current research foundation of flipped classrooms. For instance, in class, teachers can use problem-based learning approaches to guide students to develop their problem-solving ability in combination with the development of their mental processes (Lin & Lin, 2016).

Acknowledgement

This study is supported by the National Science and Technology Council, Taiwan, under grants NSTC 111-2410-H-153-012, NSTC 110-2511-H-153-002-MY3, and NSTC 108-2511-H-153-006-MY2. This study is also supported by the Ministry of Education, Taiwan, R.O.C. under grants MOE PEE1110245.

References

Abdullah, M. Y., Hussin, S. & Ismail, K. (2019). Investigating the effects of the flipped classroom model on Omani EFL learners' motivation level in English speaking performance. *Education and Information Technologies, 24*, 2975-2995. https://doi.org/10.1007/s10639-019-09911-5

Akçayır, G., & Akçayır, M. (2018). The Flipped classroom: A Review of its advantages and challenges. *Computers & Education, 126*, 334-345. https://doi.org/10.1016/j.compedu.2018.07.021

Al-Ahdal, A. A. M. H. (2020). EBook interaction logs as a tool in predicting learner performance in reading. *Asiatic: IIUM Journal of English Language and Literature, 14*(1), 174-188.

Alamri, M. M. (2019). Students' academic achievement performance and satisfaction in a flipped classroom in Saudi Arabia. *International Journal of Technology Enhanced Learning, 11*(1), 103-119. https://doi.org/10.1504/IJTEL.2019.096786

Baker, A., Navarro, E. O., & van der Hoek, A. (2005). An Experimental card game for teaching software engineering processes. *Journal of Systems and Software, 75*(1-2), 3-16. https://doi.org/10.1016/j.jss.2004.02.033

Barkley, E. (2010). *Student engagement techniques: A Handbook for college faculty*. Jossey-Bass.

Bergmann, J., & Sams, A. (2012). *Flip your classroom: Reach every student in every class every day*. International Society for Technology in Education.

Birgili, B., Seggie, F. N., & Oğuz, E. (2021). The Trends and outcomes of flipped learning research between 2012 and 2018: A Descriptive content analysis. *Journal of Computers in Education, 8*, 365-394. https://doi.org/10.1007/s40692-021-00183-y

Bond, M. (2020). Facilitating student engagement through the flipped learning approach in K-12: A Systematic review. *Computers & Education, 151*, 103819. https://doi.org/10.1016/j.compedu.2020.103819

Boticki, I., Akçapınar, G., & Ogata, H. (2019). E-book user modelling through learning analytics: The Case of learner engagement and reading styles. *Interactive Learning Environments, 27*(5-6), 754-765. https://doi.org/10.1080/10494820.2019.1610459

Bryson, C., & Hand, L. (2007). The Role of engagement in inspiring teaching and learning. *Innovations in Education and Teaching International, 44*(4), 349-362. https://doi.org/10.1080/14703290701602748

Cai, H., & Gu, X. (2022). Factors that influence the different levels of individuals' understanding after collaborative problem solving: the effects of shared representational guidance and prior knowledge. *Interactive Learning Environments, 30*(4), 695-706. https://doi.org/10.1080/10494820.2019.1679841

Chang, S. C., & Hwang, G. J. (2018). Impacts of an augmented reality-based flipped learning guiding approach on students' scientific project performance and perceptions. *Computers & Education, 125*, 226-239. https://doi.org/10.1016/j.compedu.2018.06.007

Chen, C. H., & Su, C. Y. (2019). Using the BookRoll e-book system to promote self-regulated learning, self-efficacy and academic achievement for university students. *Educational Technology & Society, 22*(4), 33-46.

Chung, C. J., Lai, C. L., & Hwang, G. J. (2021). Roles and research trends of flipped classrooms in nursing education: a review of academic publications from 2010 to 2017. *Interactive Learning Environments, 29*(6), 883-904. https://doi.org/10.1080/10494820.2019.1619589

Cico, O., Jaccheri, L., Nguyen-Duc, A., & Zhang, H. (2021). Exploring the intersection between software industry and software engineering education-A Systematic mapping of Software Engineering Trends. *Journal of Systems and Software, 172*, 110736. https://doi.org/10.1016/j.jss.2020.110736

Coates, H. (2006). *Student engagement in campus-based and online education: University connections*. Routledge.

Connor, C. M., Day, S. L., Zargar, E., Wood, T. S., Taylor, K. S., Jones, M. R., & Hwang, J. K. (2019). Building word knowledge, learning strategies, and metacognition with the Word-Knowledge e-Book. *Computers & Education*, *128*, 284-311. https://doi.org/10.1016/j.compedu.2018.09.016

Dembedza, M.P., & Chipurura, B. (2020). Evaluating experiences and perceptions of undergraduate Zimbabwean nutrition students on flipped-classroom learning: A Mixed methods approach. *IOSR Journal of Research & Method in Education, 10*(2), 1-7.

DeRuisseau, L. R. (2016). The Flipped classroom allows for more class time devoted to critical thinking. *Advances in Physiology Education*, *40*(4), 522–528. https://doi.org/10.1152/advan.00033.2016

Erturan Ilker, G., Arslan, Y., & Demirhan, G. (2014). A Validity and reliability study of the motivated strategies for learning questionnaire. *Educational Sciences: Theory and Practice*, *14*(3), 829-833.

Esperanza, P. J., Himang, C., Bongo, M., Selerio Jr., E., & Ocampo, L. (2021). The Utility of a flipped classroom in secondary Mathematics education. *International Journal of Mathematical Education in Science and Technology.* https://doi.org/10.1080/0020739X.2021.1957166

Fahmi, R., Friatin, L., & Irianti, L. (2020). The Use of flipped classroom model in reading comprehension. *Journal of Applied Linguistics and Literacy, 4*(1), 77-94.

Garousi, V., Giray, G., Tuzun, E., Catal, C., & Felderer, M. (2020). Closing the gap between software engineering education and industrial needs. *IEEE Software, 37*(2), 68-77. https://doi.org/10.1109/MS.2018.2880823

Gilboy, M. B., Heinerichs, S., & Pazzaglia, G. (2015). Enhancing student engagement using the flipped classroom. *Journal of Nutrition Education and Behavior, 47*(1), 109-114. https://doi.org/10.1016/j.jneb.2014.08.008

Haghighi, H., Jafarigohar, M., Khoshsima, H., & Vahdany, F. (2019). Impact of flipped classroom on EFL learners' appropriate use of refusal: achievement, participation, perception. *Computer Assisted Language Learning, 32*(3), 261-293. https://doi.org/10.1080/09588221.2018.1504083

Horn, M. B., & Staker, H. (2011). *The Rise of K-12 blended learning.* Innosight Institute.

Huang, A. Y. Q., Lu, O. H. T., & Yang, S. J. H. (2023). Effects of artificial Intelligence–Enabled personalized recommendations on learners' learning engagement, motivation, and outcomes in a flipped classroom. *Computers & Education, 194,* 104684. https://doi.org/10.1016/j.compedu.2022.104684

Huang, A. Y. Q., Lu, O. H. T., Huang, J. C. H., Yin, C. J., & Yang, S. J. H. (2020). Predicting students' academic performance by using educational big data and learning analytics: evaluation of classification methods and learning logs. *Interactive Learning Environments, 28*(2), 206-230. https://doi.org/10.1080/10494820.2019.1636086

Huang, H., & Hwang, G. J. (2019). Facilitating inpatients' family members to learn: A Learning engagement-promoting model to develop interactive E-Book systems for patient education. *Educational Technology & Society, 22* (3), 74–87.

Huang, Y. M., Silitonga, L. M., & Wu, T. T. (2022). Applying a business simulation game in a flipped classroom to enhance engagement, learning achievement, and higher-order thinking skills. *Computers & Education, 183*, 104494. https://doi.org/10.1016/j.compedu.2022.104494

Hwang, G. J., & Chang, C. Y. (2020). Facilitating decision-making performances in nursing treatments: A Contextual digital game-based flipped learning approach. *Interactive Learning Environments*. https://doi.org/10.1080/10494820.2020.1765391

Hwang, G. J., & Chang, H. F. (2011). A Formative assessment-based mobile learning approach to improving the learning attitudes and achievements of students. *Computers & Education, 56*(4), 1023-1031. https://doi.org/10.1016/j.compedu.2010.12.002

Hwang, G. J., & Lai, C. L. (2017). Facilitating and bridging out-of-class and in-class learning: An Interactive e-book-based flipped learning approach for math courses. *Journal of Educational Technology & Society*, *20*(1), 184-197.

Kurnianto, B., Wiyanto, W., & Haryani, S. (2019). Critical thinking skills and learning outcomes by improving motivation in the model of flipped classroom. *Journal of Primary Education, 8*(6), 282-291.

Lai, H. M., Hsieh, P. J., Uden, L., & Yang, C. H. (2021). A Multilevel investigation of factors influencing university students' behavioral engagement in flipped classrooms. *Computers & Education, 175,* 104318. https://doi.org/10.1016/j.compedu.2021.104318

Lee, G. G., Jeon, Y. E., & Hong, H. G. (2021). The Effects of cooperative flipped learning on science achievement and motivation in high school students. *International Journal of Science Education, 43*(9), 1381-1407. https://doi.org/10.1080/09500693.2021.1917788

Lew, E. K. (2016). Creating a contemporary clerkship curriculum: The Flipped classroom model in emergency medicine. *International journal of emergency medicine, 9*(1), 25. https://doi.org/10.1186/s12245-016-0123-6

Lim, J., Whitehead, G. E. K., & Choi, Y. (2021). Interactive e-book reading vs. paper-based reading: Comparing the effects of different mediums on middle school students' reading comprehension. *System*, *97*, 102434. https://doi.org/10.1016/j.system.2020.102434

Lin, Y. T. (2016). When mobile technology meets traditional classroom learning environment: How does it improve students' learning performances? In K. Wallace (Ed.), *Learning environments: Emerging theories, applications and future directions (pp. 143-158)*. Nova Science Publishers, Inc.

Lin, Y. T. (2019). Impacts of a flipped classroom with a smart learning diagnosis system on students' learning performance, perception, and problem solving ability in a software engineering course. *Computers in Human Behavior, 95*, 187-196. https://doi.org/10.1016/j.chb.2018.11.036

Lin, Y. T. (2021). Effects of flipped learning approaches on students' learning performance in software engineering education. *Sustainability, 13*(17), 9849. https://doi.org/10.3390/su13179849

Lin, Y. T., & Cheng, C. T. (2022). Effects of Technology-Enhanced Board Game in Primary Mathematics Education on Students' Learning Performance. *Applied Sciences*, *12*(22), 11356. https://doi.org/10.3390/app122211356

Lin, Y. T., & Lin, Y. C. (2016). Effects of mental process integrated nursing training using mobile device on students' cognitive load, learning attitudes, acceptance, and achievements. *Computers in Human Behavior*, *55*, 1213-1221. https://doi.org/10.1016/j.chb.2015.03.076

Lin, Y. T., Lin, Y. C., Huang, Y. M., & Cheng, S. C. (2013). A Wiki-based teaching material development environment with enhanced particle swarm optimization. *Educational Technology & Society*, *16*(2), 103-118.

Lin, Y. T., Tseng, Y. M., Lee, Y. S., Wang, T. C., Tsai, S. I., & Yi, Y. J. (2018). Development of a SoLoMo game-based application for supporting local cultural learning in Taiwan. *Educational Technology & Society*, *21*(4), 115-128.

Liu, C., Wan, P., Tu, Y. F., Chen, K., & Wang, Y. (2021). A WSQ-based mobile peer assessment approach to enhancing university students' vocal music skills and learning perceptions. *Australasian Journal of Educational Technology*, *37*(6), 1-17. https://doi.org/10.14742/ajet.6832

Lo, C. K., & Hew, K. F. (2017). Using "first principles of instruction" to design secondary school mathematics flipped classroom: The Findings of two exploratory studies. *Educational Technology & Society*, *20*(1), 222-236.

Lo, C. K., Hew, K. F., & Chen, G. (2017). Toward a set of design principles for mathematics flipped classrooms: A Synthesis of research in mathematics education. *Educational Research Review*, *22*, 50-73. https://doi.org/10.1016/j.edurev.2017.08.002

Malik, S. I., Mathew, R., Al-Sideiri, A., Jabbar, J., Al-Nuaimi, R., & Tawafak, R. M. (2022). Enhancing problem-solving skills of novice programmers in an introductory programming course. *Computer Applications in Engineering Education, 30*(1), 174-194. https://doi.org/10.1002/cae.22450

Mamun, M. A. A., Lawrie, G., & Wright, T. (2020). Instructional design of scaffolded online learning modules for self-directed and inquiry-based learning environments. *Computers & Education, 144*, 103695. https://doi.org/10.1016/j.compedu.2019.103695

Martínez-Jiménez, R., & Ruiz-Jiménez, M. C. (2020). Improving students' satisfaction and learning performance using flipped classroom. *The International Journal of Management Education, 18*(3), 100422. https://doi.org/10.1016/j.ijme.2020.100422

Matturro, G., Raschetti, F., & Fontán, C. (2019). A Systematic mapping study on soft skills in software engineering. *Journal of Universal Computer Science, 25*(1), 16-41. https://doi.org/10.3217/jucs-025-01-0016

McNally, B., Chipperfield, J., Dorsett, P., Del Fabbro, L., Frommolt, V., Goetz, S., Lewohl, J., Molineux, M., Pearson, A., Reddan, G., Roiko, A., & Rung, A. (2017). Flipped classroom experiences: student preferences and flip strategy in a higher education context. *Higher Education*, *73*(2), 281-298. https://doi.org/10.1007/s10734-016-0014-z

Merkle, A. C., Ferrell, L. K., Ferrell, O. C., & Hair, J. F. (2022). Evaluating e-book effectiveness and the impact on student engagement. *Journal of Marketing Education, 44*(1), 54-71. https://doi.org/10.1177/02734753211035162

Mouri, K., Uosaki, N., & Ogata, H. (2018). Learning analytics for supporting seamless language learning using e-book with ubiquitous learning system. *Educational Technology & Society*, *21*(2), 150-163.

Mukhlisa, R., Gani, A., Winarni, S., Khaldun, I., & Hanum, L. (2021). Independence of learning and achievement of learners' cognitive abilities in thermochemical materials through the application of flipped classroom. *Jurnal Penelitian Pendidikan IPA, 7*(4), 523-530. https://doi.org/10.29303/jppipa.v7i4.674

Ng, B. (2018). The Neuroscience of growth mindset and intrinsic motivation. *Brain Sciences*, *8*(2), 20. https://doi.org/10.3390/brainsci8020020

Ni'mah, D., & Umamah, A. (2020). Mobile-assisted on extensive reading: students' voices. *English Education: Journal of English Teaching and Research, 5*(1), 41-48. https://doi.org/10.29407/jetar.v5i1.14367

Ogata, H., Oi, M., Mohri, K., Okubo, F., Shimada, A., Yamada, M., Wang, J., & Hirokawa, S. (2017). Learning analytics for ebook-based educational big data in higher education. In H. Yasuura, C. Kyung, Y. Liu, & Y. Lin (Eds.), *Smart Sensors at the IoT Frontier* (pp. 327-350). Springer.

Ogata, H., Yin, C., Oi, M., Okubo, F., Shimada, A., Kojima, K., & Yamada, M. (2015). e-Book-based learning analytics in university education. In H. Ogata, W. Chen, S. C. Kong, & F. Qiu (Eds.), *Proceedings of the 23rd International Conference on Computers in Education* (pp. 401-406). Asia-Pacific Society for Computers in Education.

Oweis, T. I. (2018). Effects of using a blended learning method on students' achievement and motivation to learn English in Jordan: A Pilot case study. *Education Research International, 2018*. https://doi.org/10.1155/2018/7425924

Palinussa, A. L., Laamena, C., & Mananggel, M. B. (2021). Development of learning models flipped classroom to increase the mathematical Higher Order Thinking Skills (HOTS) of students in high school. *Technology Reports of Kansai University, 63*(3), 7349-7355.

Patanwala, A. E., Erstad, B. L., & Murphy, J. E. (2017). Student use of flipped classroom videos in a therapeutics course. *Currents in Pharmacy Teaching and Learning*, *9*(1), 50-54. https://doi.org/10.1016/j.cptl.2016.08.043

Pintrich, P. R., & De Groot, E. V. (1990). Motivational and self-regulated learning components of classroom academic performance. *Journal of Educational Psychology, 82*(1), 33-40. https://doi.org/10.1037/0022-0663.82.1.33

Rasheed, R. A., Kamsin, A., & Abdullah, N. A. (2020). Challenges in the online component of blended learning: A Systematic review. *Computers & Education, 144*, 103701. https://doi.org/10.1016/j.compedu.2019.103701

Roehl, A., Reddy, S. L., & Shannon, G. J. (2013). The Flipped classroom: An Opportunity to engage millennial students through active learning strategies. *Journal of Family & Consumer Sciences*, *105*(2), 44-49.

Ryan, R. M., & Deci, E. L. (2019). Chapter Four - Brick by brick: The Origins, development, and future of self-determination theory. In A. J. Elliot (Ed.), *Advances in Motivation Science* (Vol. 6, pp. 111-156). Elsevier. https://doi.org/10.1016/bs.adms.2019.01.001

Shanta, S., & Wells, J. G. (2022). T/E design based learning: assessing student critical thinking and problem solving abilities. *International Journal of Technology and Design Education, 32*, 267-285. https://doi.org/10.1007/s10798-020-09608-8

Siswanto, R. A. (2021). The Implementation of flipped classroom in improving students' writing skills at the University of Brawijaya. *Jurnal Ilmu Pendidikan, 27*(2), 58-63.

Sung, H. Y., Hwang, G. J., Chen, C. Y., & Liu, W. X. (2022). A Contextual learning model for developing interactive e-books to improve students' performances of learning the Analects of Confucius. *Interactive Learning Environments, 30*(3), 470-483. https://doi.org/10.1080/10494820.2019.1664595

Sutcliffe, A., & Hart, J. (2016). Analyzing the role of interactivity in user experience. *International Journal of Human–Computer Interaction*, *33*(3), 229-240. https://doi.org/10.1080/10447318.2016.1239797

Svenningsson, J., Höst, G., Hultén, M. & Hallström, J. (2022). Students' attitudes toward technology: Exploring the relationship among affective, cognitive and behavioral components of the attitude construct. *International Journal of Technology and Design Education, 32,* 1531-1551. https://doi.org/10.1007/s10798-021-09657-7

van Alten, D. C. D., Phielix, C., Janssen, J., & Kester, L. (2020). Self-regulated learning support in flipped learning videos enhances learning outcomes. *Computers & Education, 158*, 104000. https://doi.org/10.1016/j.compedu.2020.104000

Wu, T. T. & Wu, Y. T. (2020). Applying project-based learning and SCAMPER teaching strategies in engineering education to explore the influence of creativity on cognition, personal motivation, and personality traits. *Thinking Skills and Creativity, 35*, 100631. https://doi.org/10.1016/j.tsc.2020.100631

Yang, C. C. Y., Chen, I. Y. L., Akçapınar, G., Flanagan, B., & Ogata, H. (2021). using a summarized lecture material recommendation system to enhance students' preclass preparation in a flipped classroom. *Educational Technology & Society, 24*(2), 1-13.

Ye, X. D., Chang, Y. H., & Lai, C. L. (2019). An Interactive problem-posing guiding approach to bridging and facilitating pre- and in-class learning for flipped classrooms. *Interactive Learning Environments, 27*(8), 1075-1092. https://doi.org/10.1080/10494820.2018.1495651

Yin, C., Yamada, M., Oi, M., Shimada, A., Okubo, F., Kojima, K., & Ogata, H. (2019). Exploring the relationships between reading behavior patterns and learning outcomes based on log data from e-books: A Human Factor Approach. *International Journal of Human–Computer Interaction*, *35*(4-5), 313-322. https://doi.org/10.1080/10447318.2018.1543077

Zengin, Y. (2017). Investigating the use of the Khan Academy and mathematics software with a flipped classroom approach in mathematics teaching. *Educational Technology & Society*, *20*(2), 89-100.

Zhao, J., Hwang, G. J., Chang, S. C., Yang, Q. F., & Nokkaew, A. (2021). Effects of gamified interactive e-books on students' flipped learning performance, motivation, and meta-cognition tendency in a mathematics course. *Educational Technology Research and Development, 69*, 3255-3280. https://doi.org/10.1007/s11423-021-10053-0

Zou, D. (2020). Gamified flipped EFL classroom for primary education: student and teacher perceptions. *Journal of Computers in Education*, *7*, 213-228. https://doi.org/10.1007/s40692-020-00153-w