Chen, B., Hwang, G.-H., & Wang, S.-H. (2021). Gender Differences in Cognitive Load when Applying Game-Based Learning with Intelligent Robots. *Educational Technology & Society*, *24* (3), 102–115.

Gender Differences in Cognitive Load when Applying Game-Based Learning with Intelligent Robots

Beyin Chen[1], Gwo-Haur Hwang[2*] and Shen-Hua Wang[3]
[1]Department of Information Technology, Ling Tung University, Taiwan // [2]Bachelor Program in Industrial Technology, National Yunlin University of Science and Technology, Taiwan // [3]Department of Information Management, Ling Tung University, Taiwan // byc@teamail.ltu.edu.tw // ghhwang0424@gmail.com // happy0487587@gmail.com
[*]Corresponding author

ABSTRACT: The application of artificial intelligence (AI) in education is now widespread, and the use of robots in education has demonstrated a positive influence on students' behavior and development. However, the use of emerging technologies usually results in cognitive load, especially for elementary school students whose learning capacity has not yet been established. In addition, students of different genders have different physical, psychological and learning characteristics, so gender differences affect cognitive load. Cognitive load can be divided into two types: positive cognitive load and negative cognitive load. Usually, positive cognitive load results in good learning performance while negative cognitive load results in bad learning performance. Therefore, we use the cognitive load theory to define learning efficiency as the co-impact of learning performance and cognitive load. We take game-based intelligent robots for Chinese idiom learning as an example, and explore the impacts of gender differences on elementary school students. To achieve these aims, this study combined games and Zenbo robots, and applied them to educate elementary school students in the use of Chinese idioms. Secondly, this study conducted an experiment and analyzed the experimental results. The participants were 24 fourth-grade elementary school students from the central region of Taiwan. Results showed that this system is more beneficial for boys as their cognitive load was significantly lower. Boys' learning performance was also better, although the difference did not reach significance. Furthermore, learning efficiency for boys was significantly higher. Reasons for these results are explained.

Keywords: Artificial intelligence, Cognitive load theory, Game-based learning, Gender differences, Robots

1. Introduction

The work of artificial intelligence (AI) is dedicated to solving cognitive problems that are usually related to human intelligence. There is no denying that the education sector has been significantly affected by AI. AI in education (AIEd) is now widely used by learners and educators (Chen, Xie, & Hwang, 2020; Hwang, Xie, Wah, & Gašević, 2020). For example, Hwang, Sung, Chang, and Huang (2020) developed an adaptive learning system and explored its associated mathematical anxiety and cognitive load. Their experimental results showed that the proposed approach helped the low achievers successfully complete the learning tasks. A form of AI in education is the use of social robots (Papadopoulos, Lazzarino, Miah, & Weaver, 2020). Papadopoulos et al. (2020) pointed out that the use of robots in education has demonstrated a positive influence on the behavior and development of students, especially in the areas of problem-solving skills (Barak & Zadok, 2009) and teamwork practice (Varney, Janoudi, Aslam, & Graham, 2012), increased motivation to learn (Kubilinskiene, Zilinskiene, Dagiene, & Sinkevicius, 2017), and enhancement of participation (Rusk, Resnick, Berg, & Pezalla-Granlund, 2008).

Microsoft founder and chairman Bill Gates predicted that the robot industry will become the next hot area (Gates, 2007). Due to the rapid development of the future robot industry, the urgency of promoting robot education or related cross-disciplinary education is expected to increase day by day (Hsiao & Huang, 2012). Westlund et al. (2017) pointed out that robots leverage human means of communicating, such as speech, movement and nonverbal cues, including gaze, gestures, and facial expressions, in order to interface with us in more natural ways. Their study also pointed out that the emotional expressiveness of the robot's speech might modulate children's learning. Therefore, leveraging robots to help elementary school students learn idioms is a feasible solution.

On the other hand, the theoretical basis of the influence of games on learning is derived from children's cognitive development psychology. Cognitive development theory states that the growth of learners in different cognitive stages needs to be completed by the maturity and transformation of the previous stage (Piaget, 1964). Cai, Yan, Yang, and Wang (2012) also pointed out that if game situations can be used to help children to learn joyfully, the children's focus quality during the learning process can last longer, and then the stage's maturity and transformation can be reached.

In the past, several studies also integrated games and robots. For example, the robot Mindstorms integrates system simulation and program manipulation into games. It creates a broad field of vision for learning to help users learn in more depth, and has become a classic example of joyful learning (Resnick & Ocko, 1991; Rusk et al., 2008). Liu and Lin (2009) also pointed out that the impacts of combining games and smart robots will be different from those of traditional simple game-based learning, which creates a broader research area for joyful learning.

Emerging technologies and game-based learning can help students increase their learning interest (Ng'ambi, 2013). However, they usually result in cognitive load, especially for younger students whose learning capacity has not yet been established. For emerging technologies, Zhong, Zheng, and Zhan (2020) examined the effects of virtual and physical robots (VPR) used in different learning stages (simple session/complex session) in a robotics programming course. They found that significant difference existed in engineering design ability and cognitive load, no matter whether in simple or complex learning sessions. Huang, Shadiev, and Hwang (2016) explored the effectiveness of applying speech-to-text recognition (STR) technology during lectures in English on the cognitive load of non-native English speaking students. The result showed that lectures in English caused less cognitive load for low ability EFL students when they used STR-texts. For game-based learning, Liao, Chen, and Shih (2019) investigated how the use of an instructional video and collaboration influenced the intrinsic motivation and cognitive load of students learning Newtonian mechanics within a digital game-based learning (DGBL) environment. While collaborative DGBL promoted intrinsic motivation, the results for cognitive load showed that the use of an instructional video in collaborative DGBL significantly reduced both intrinsic and extraneous cognitive loads. Javora, Hannemann, Stárková, Volná, and Brom (2019) examined the effects of a holistic and appealing visual design of a learning game. Based on cognitive-affective theory of learning with media and cognitive load theory, they found that visual design's influence on learning outcomes is mediated by (at least) two hidden variables: cognitive engagement and cognitive load.

However, Bevilacqua (2017) pointed out that gender differences in cognitive load have resulted in some differences in aspects of associated working memory systems that are relevant to cognitive load theory. He also indicated that if males and females process information differently in working memory, cognitive load levels will be different for males and females experiencing similar stimuli under certain conditions. Many studies have pointed out that in different learning environments, gender may have an impact on cognitive load. For example, Christophel and Schnotz (2017) explored the correlations between cognitive load and competences. They found that the correlations differed for female and male participants. Wong, Castro-Alonso, Ayres, and Paas (2015) explored gender effects when learning manipulative tasks from instructional animations and static presentations. They found that the cognitive load of females and males have reverse results.

However, in our survey of the literature, we found that past studies exploring gender differences in cognitive load seldom investigated robots. Therefore, our study explored the impacts of gender differences in cognitive load on applying game-based learning by intelligent robots. Sweller (1998) defined cognitive load as the amount of load generated when a specific task is applied to an individual's cognitive system. Its management is embodied within the framework of the cognitive load theory (CLT) (Sweller, 2005). CLT has been proven to be a theory of great value for instructional design (Paas, van Gog, & Sweller, 2010). If the measurement of cognitive load is combined into the related aspects of CLT, it is likely to be more accurate and complete. The related aspects include cognitive load, learning performance and learning efficiency (Paas & van Merriënboer, 1993). Among these three aspects, learning efficiency reflects the co-impact of learning performance and cognitive load. The goal of our study is that when the educational robots are used by elementary school students, learning efficiency can be better exerted in response to the influence of gender differences.

Based on the above, the research questions of this study are described as follows:
RQ1: Does gender have a significant impact on the cognitive load of elementary school students in learning idioms?
RQ2: Does gender have a significant impact on the learning performance of elementary school students in learning idioms?
RQ3: Does gender have a significant impact on the learning efficiency of elementary school students in learning idioms?

2. Literature review

2.1. Artificial intelligence (AI) in education

The rapid development of computing and information processing technology has accelerated the development and application of AI, which aims to enable computers to perform tasks via simulating human intelligent behaviors, such as reasoning, analysis, and decision-making (Hwang, Xie, Wah, & Gašević, 2020). From the perspective of precision education, it emphasizes the need to provide prevention and intervention measures for learners by analyzing their learning conditions or behaviors (Hart, 2016). Chen, Xie, and Hwang (2020) indicated that AIEd is widely used by learners and educators nowadays, and involves various tools and applications, for example, intelligent tutoring systems, teaching robots, and adaptive learning systems. AIEd supports learning in traditional classes and workplaces by combining AI with various learning sciences such as education, psychology, linguistics and neuroscience, and aims to stimulate and promote AI-driven educational application (Luckin, Holmes, Griffiths & Forcier, 2016).

With the development of AI technology, the "natural language processing (NLP)" feature of robots is playing a pivotal role. NLP is the ability of computers and cloud-based applications (apps) to communicate with humans in their own natural languages, such as Chinese or English (Smith, Haworth, & Žitnik, 2020). It refers to building computational tools that analyze and represent human language at human communication complexity levels (Liddy, 2001). Several innovative educational products have claimed their adoption of AI-enabled techniques to facilitate learning performance, with applications ranging from chatbots with NLP techniques (Crossley, Allen, Snow, & McNamara, 2016). In our research, we used robots' NLP feature to interact with elementary school students to learn Chinese idioms.

2.2. Application of robots in elementary school education

Emerging technologies are transforming society and inspiring technological innovation in previously un-thought-of practices, beliefs and perceptions (Ng'ambi, 2013). Robots are one of the increasingly popular emerging technologies that have the potential to influence students' learning (Kucuk & Sisman, 2017). Several studies have reported that the use of robots helps children engage more in their learning activities. For example, Wu, Wang, and Chen (2015) used the robot's facial expressions, gestures, and movements to generate various forms of communication and interaction with the students, thereby helping the elementary school students learn English. The experimental results showed that the students' learning experience, motivation and achievement improved. Hsiao et al. (2019) used robot-based practices to develop an activity that incorporated the 6E (engage, explore, explain, engineer, enrich, and evaluate) model to improve elementary school students' learning effects. With the 6E model, the instructor facilitated the students' hand-made process by strengthening the connection between life experience, learning content, learning characterization, and interdisciplinary knowledge.

However, in the past, some studies indicated that gender may affect the use of educational robots for elementary school students. For example, Master, Cheryan, Moscatelli, and Meltzoff (2017) indicated that young girls had less interest and self-efficacy in technology compared with boys in elementary school. Therefore, considering the issue of gender differences in cognitive load is necessary.

2.3. Cognitive load theory (CLT) and the impacts of gender on cognitive load

Cognitive load (CL) in e-learning has been explored for many years and its management is embodied within the framework of the CLT (Sweller, 2005). The CLT emphasizes that all novel information is initially processed by working memory which has capacity and duration limitations; the information is then stored in long-term memory which is unlimited. The aim of instructional design should be to reduce unnecessary working memory loads, and free the capacity for learning-related processing to accommodate the limited capacity of working memory (Mutlu-Bayraktar, Cosgun, & Altan, 2019). CL has two meanings. One is a causal dimension that reflects the interaction between personal traits and task traits. The other is a dimension of measurement that describes the measurable aspects of mental load (ML), mental effort (ME), and performance (PE) (Krell, 2017; Paas & van Merriënboer, 1994). ML is task-related and refers to the cognitive ability required to handle task complexity. Conversely, ME is subject-related and reflects the cognitive ability invested when a person is working on a task. The relationship between PE and CL is unknown. For example, a subject may have the same number of correct answers (i.e., PE) in one test, but may require a different amount of ME (Paas, Tuovinen, Tabbers, & van Gerven, 2003). Paas and van Merriënboer (1993) proposed a method that provides a tool to correlate ME measurements of PE. This approach defines high learning efficiency as low ME to achieve high PE, while low learning efficiency is defined as high ME to achieve low PE.

However, gender differences may result in differences in cognitive load (Bevilacqua, 2017). In the past, many researchers explored gender difference in cognitive load. For example, Wong, Castro-Alonso, Ayres, and Paas (2015) explored gender effects when learning manipulative tasks from instructional animations and static presentations. The results showed that for females, on such tasks, using animations might have clearer advantages in managing their cognitive load rather than static presentations. For males, the reverse strategy may be more effective. Hwang, Hong, Cheng, Peng, and Wu (2013b) conducted experiments to explore the gender differences in cognitive load for sixth-grade students. The results showed that girls had a higher cognitive load and more competition anxiety from synchronous types of competitive games.

Although many studies have explored the influence of gender on cognitive load in various learning environments in the past, we have rarely found research on the influence of gender in cognitive load in the learning environment of educational robots. Therefore, we believe that it is a worthwhile topic to explore the impact of gender on the cognitive load of elementary school students in the learning environment of educational robots. These aspects include cognitive load (CL), learning performance (PE) and learning efficiency (LE). We measured cognitive load (CL) by defining the questions of the CL questionnaire as the sum of the questions of the mental load (ML) questionnaire and the mental effort (ME) questionnaire (Hwang, Yang, & Wang, 2013a). It was expected that the application of CLT could make the issue of the impact of gender difference on cognitive load better understood from a wider range of perspectives.

3. System development

3.1. System architecture

In this study, we combined intelligent robots and game-based learning to enhance students' learning. Among various types of robots, we chose Zenbo, which was launched by ASUS (i.e., ASUSTeK Computer Inc. in Taiwan) in January 2017. It is 62 cm tall and weighs 10 kg. It can display 24 expressions such as happiness, shyness and vitality, as well as body movements such as raising its head, walking and rotation (see https://www.youtube.com/watch? v=zW23nlYRWCk). The shape, sound and body movements are all designed to simulate the cute image of a two-year-old boy. Zenbo is a type of robot with the NLP feature, which is one of the important technologies of modern AI. These characteristics make Zenbo highly suitable for elementary school students (Chen, Hwang, Wang, & Peng, 2018). Based on these, we developed the "Zenbo Idiom Learning System (ZILS)" to help elementary school students learn idioms. This system is divided into two modules. They are the "idiom learning" module and the "reviewing with games" module. A unit includes seven idioms and takes 40 minutes to complete. The system architecture is shown in Figure 1.

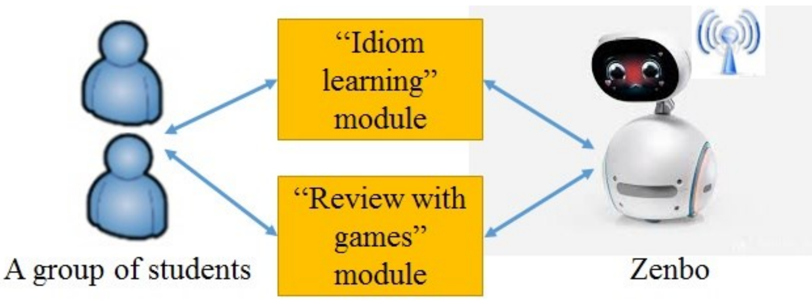

Figure 1. System architecture

3.2. The "idiom learning" module

Since idioms are defined as linguistic expressions, their overall meaning cannot be predicted from the meanings of the constituent parts (literal meaning) (Kovecses & Szabco, 1996). Therefore, idiom interpretation is important for elementary school students. In addition, the Chinese poet Yu (1994) stated that idioms are a unique feature of the Chinese language, often with historical stories and philosophical significance. Yu (1994) pointed out that, at present, many people who write Chinese do not use idioms. He also stated that the decline in the use of idioms shows that classical Chinese is being forgotten and its cultural significance is shrinking. To arouse students’ interest in using idioms, besides understanding their meaning, the classical allusions of idioms and idiom sentence-making are important.

Based on these reasons, the “idiom learning” module mainly teaches idiom interpretation, the classical allusion of idioms, and idiom sentence-making. In the “idiom learning” module, the content of the seven idioms is displayed and played by Zenbo’s screen and voice. The process is interspersed with pictures and texts. We chose the idioms according to themes such as numbers, animals, colors, people, etc. Each unit deals with a different theme, for example: for the first unit, we selected animals as the theme and then we selected seven idioms. Each idiom comprises four Chinese words. We translated the seven idioms into English as follows: “to see a bow reflected in a cup as a snake,” “the mantis stalks the cicada, unaware of the oriole behind,” “the tortoise and the hare,” “to mend the pen after the sheep are lost,” “like a fish back in water,” “if you ride a tiger, it’s hard to get off” and “to refrain from shooting at the rat for fear of breaking the vases.” Table 1 provides the classical allusion of the idiom of “to refrain from shooting at the rat for fear of breaking the vases.”

Table 1. An example of the classical allusions

The idiom	“to refrain from shooting at the rat for fear of breaking the vases”
The classical allusion	There is a story in the book of Han which tells of a rich man who was a lover of antiques and who had a large collection. Among them, there was a rare vase made of jade. Due to the vase having exquisite workmanship and historical value, he loved it dearly. One night, a mouse jumped into the jade vase and wanted to eat some leftovers inside. The rich man happened to see this scene. He was very annoyed. In a rage, he took a stone and smashed it on the mouse. Of course, the mouse was killed, but the precious jade vase was also broken. The loss of the vase pained the man greatly and he deeply regretted his own thoughtlessness, which brought him this unrecoverable loss. He now realized that anyone who cares for the present while overlooking consequences is apt to bring disaster upon himself. So, he warned people by saying, “Do not burn your house to get rid of a mouse.”

In the “idiom learning” module, we designed its homepage to show seven animals representing the seven idioms as shown in Figure 2. According to the students’ selection, the system enters the interpretation, the classical allusion, and the sentence-making of the idiom. Zenbo will read aloud the text of this content. This design enables a smooth combination of pictures and speech, and matches the multimedia principles proposed by Mayer (2009). In addition, Mayer (2009) pointed out that the combination of words and pictures can improve the learning outcomes for students with low prior knowledge, so this design is also helpful for low level prior knowledge children. Figure 3 shows the screen of the classical allusion of the idiom “to mend the pen after the sheep are lost.”

Figure 2. The homepage of the “idiom learning” module

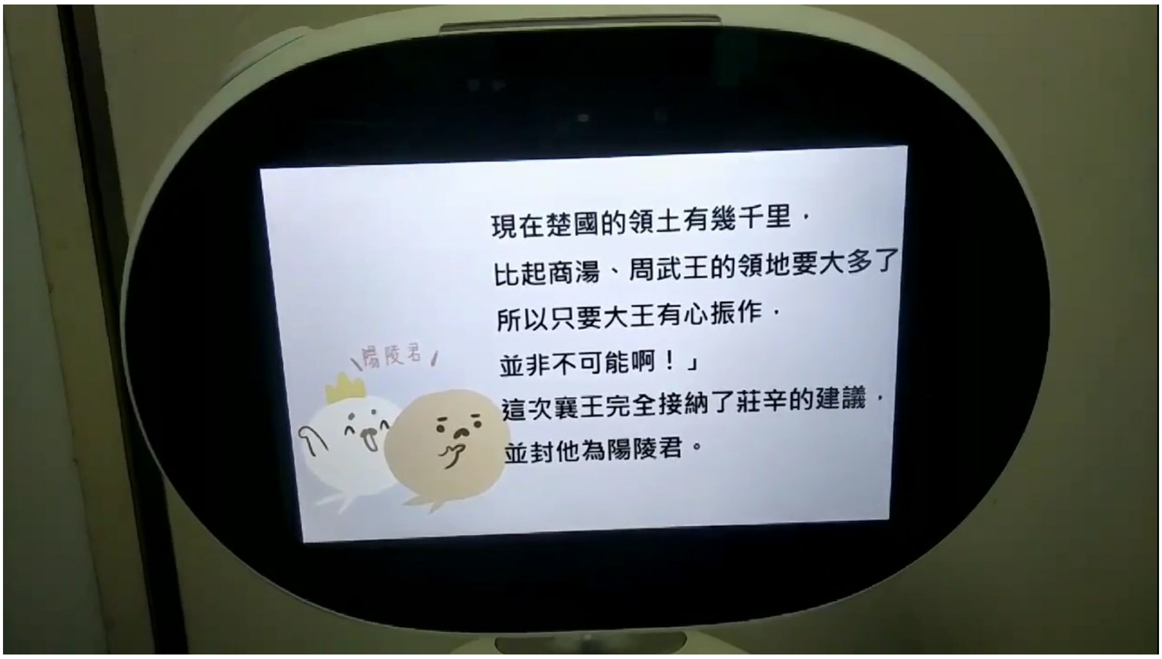

Figure 3. The classical allusion of the idiom “to mend the pen after the sheep are lost”

3.3. The “reviewing with games” module

The “reviewing with games” module reviews idioms using a game-based style. The module allows students to review the seven idioms by guessing them based on related pictures which correspond to the "words" of the idiom. Taking “to mend the pen after the sheep are lost” as an example, the first prompting picture may be a fence. The children will guess the idiom and say it out loud. If the answer is right, Zenbo will reply with a sentence such as, “Congratulations! Your answer is right.” If the answer is wrong, Zenbo will reply with a sentence such as, “Your answer is wrong. Try again!” In this case, a further prompting picture will appear. This picture may be a sheep. If the child’s answer is wrong again, Zenbo will reply with a sentence such as, “Your answer is wrong. Try for the final time!” Subsequently, the final prompting picture will appear. This picture may be a net. If the child still fails to answer the idiom completely, that idiom will be skipped. When all seven idioms have been reviewed, the failed ones will be repeated until the time is up. This “reviewing with games” module can deepen the impressions of the learned idioms.

To implement the “reviewing with games” module, we designed three prompting pictures for each idiom. Taking “to mend the pen after the sheep are lost” as an example, the three prompting pictures are a picture of a net, a picture of a sheep and a picture of a fence, as shown in Figure 4. Taking the picture of a net or a fence as an example, this is not a direct corresponding relationship with the word, but is an indirect correspondence, which can test students’ concentration ability. The embedded game elements combined with robots’ NLP function in the module “reviewing with games” facilitated students’ learning or attracted their learning attention.

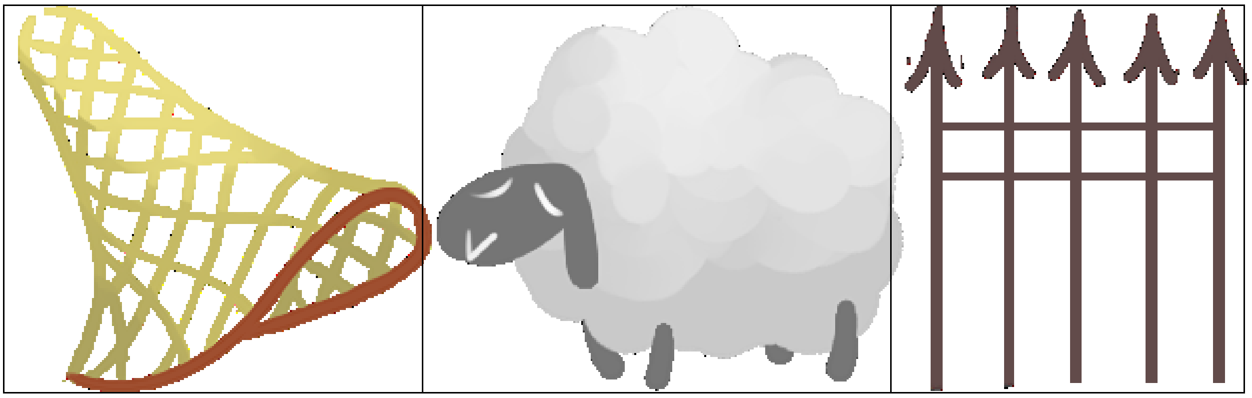

Figure 4. The three prompting pictures of the idiom “to mend the pen after the sheep are lost”

4. Research method

4.1. Research tools

In this study, the research tools include ZILS, the pre- and post-test questions, the questionnaire of cognitive load, and the SPSS statistical software. The 20 multiple-choice pre- and post-test questions were the same but in a different order. They were developed by the authors by extending the seven idioms in different aspects and expanding them into 20 questions which were compiled according to idioms appropriate for fourth-grade elementary school students in Taiwan. These 20 questions were reviewed and revised for reliability by two senior elementary school teachers, ensuring that they had expert validity. The cognitive load questionnaire was developed with reference to Hwang et al. (2013a). The questions were divided into two categories: mental load and mental effort. The Cronbach's α of the questionnaires are .92 for mental effort, .81 for mental load and .89 for cognitive load, all of which surpass the suggested threshold value of .7. These questionnaires therefore have high reliability. A 5-point Likert scale was adopted. After the experiment was finished, SPSS19 was used to analyze the data. We have selected two questions from the 20 multiple-choice questions to demonstrate the reliability of assessing students' learning performance. The questions of the cognitive load questionnaire are also listed in Table 2.

Table 2. The two selected questions and the questions of the cognitive load questionnaire

Two selected questions	1. Who is the related historical figure of the idiom "like a fish in water"? (A) Xiang Yu (B) Wen Tianxiang (C) Qu Yuan (D) Zhuge Liang 2. Which of the following statements is best used to illustrate "social laziness"? (A) The mantis catches the cicada and the oriole is behind (B) Three monks have no water to drink (C) Three people must have my teacher (D) Three days of fishing and two days of drying the net
Mental effort questions	1. The learning process of "ZILS" caused me a lot of pressure. 2. I had to put a lot of effort into completing the learning task of "ZILS". 3. The learning process of "ZILS" was difficult to understand.
Mental load questions	1. The learning content of ""ZILS" was difficult for me. 2. It took me a lot of effort to answer the "ZILS" game questions. 3. Answering "ZILS" game questions was very tiring. 4. Answering "ZILS" game questions made me feel very frustrated. 5. I didn't have enough time to answer "ZILS" game questions.

4.2. Research architecture

In this study, the independent variable is gender. All the students took advantage of one lesson to learn the seven idioms using ZILS, so the control variables are teaching time and teaching materials. The explored questions are the impacts of gender differences on the elementary school students' related aspects of CLT, including cognitive load, learning performance and learning efficiency. Therefore, the dependent variables are cognitive load, learning performance and learning efficiency. The definition of learning efficiency referred to Paas and van Merriënboer (1993). The research architecture is shown in Figure 5.

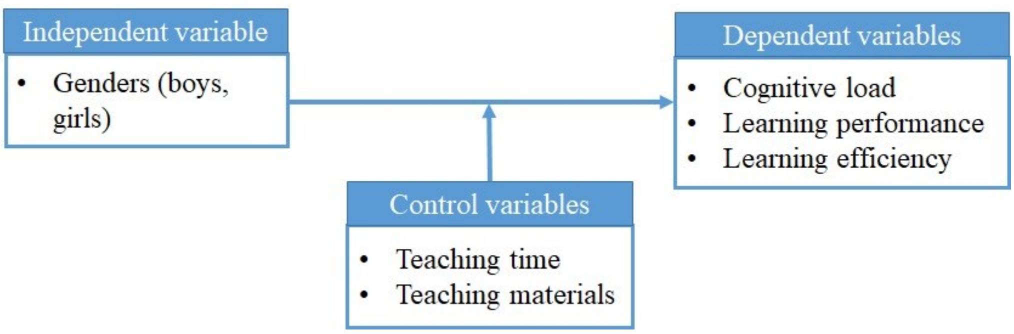

Figure 5. Research architecture

4.3. Experimental subjects

In this study, the experimental subjects are one class of fourth-grade students from an elementary school in the central region of Taiwan. There were 24 students in total including 12 boys and 12 girls. Figures 6 and 7 show scenes of the children using ZILS and filling out the questionnaires.

Figure 6. Children using ZILS

Figure 7. Children filling out questionnaires

4.4. Research flow

The research flow is shown in Figure 8. First, we introduced the system to the students and the pre-test was conducted. This stage took 25 minutes. Second, the students used ZILS to learn the seven idioms. This stage took 40 minutes which included the use of 25 minutes for the “idiom learning” module and 15 minutes for the “reviewing with games” module. During the operation of the “reviewing with games” module, the children played in groups. Each student had at least one chance to interact with Zenbo. We adopted the style of grabbing the chance to answer. When one child had answered correctly, the chance was given to another who had not yet correctly answered. If no child grabbed the chance, we assigned someone who had not passed successfully. If all the children in a group had answered successfully, they were given the chance to answer repeatedly. This learning activity process involved interaction with Zenbo’s NLP features, and included game pictures and game rules which could drive the learning atmosphere and enhance the students’ learning effectiveness. Finally, the post-test and the cognitive load questionnaire were conducted. This stage took 25 minutes.

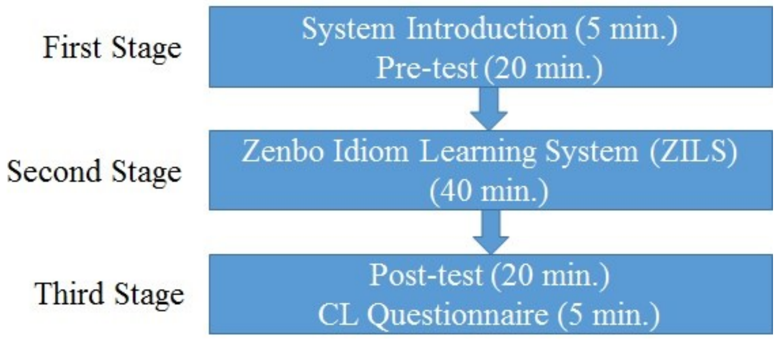

Figure 8. Research flow

5. Results

In this section, we present the results of the gender differences in the aspects of cognitive load, learning performance and learning efficiency. In the final part of this section, we explore the reasons to explain the results.

5.1. Impact on cognitive load

To explore whether there were significant differences in the three indicators of cognitive load, mental effort and mental load, we conducted the independent samples *t* test. The results are shown in Table 3 where cognitive load is the weighted average of mental load and mental effort.

Table 3. The independent samples *t* test for analyzing mental effort, mental load, and cognitive load

Load categories	Sex	Number	Average	*SD*	*t*	*Cohen’s d* (Effect size)
Mental Effort	Boys	12	1.425	0.458	2.859^{*}	0.825
	Girls	12	2.283	0.934		
Mental Load	Boys	12	1.508	0.656	2.335^{*}	0.674
	Girls	12	2.225	0.837		
Cognitive Load	Boys	12	1.464	0.506	2.844^{**}	0.872
	Girls	12	2.238	0.795		

Note. $^{*}p < .05$; $^{**}p < .01$.

From Table 3, it can be seen that boys rate significantly lower than girls for all aspects of mental effort (t = 2.859, $p < .05$, d = 0.825), mental load (t = 2.335, $p < .05$, d = 0.674) and cognitive load (t = 2.844, $p < .05$, d = 0.872). Cohen (1988) has provided benchmarks to define small (d = 0.2), medium (d = 0.5), and large (d = 0.8) effects. Accordingly, the effect sizes of the aforementioned results are all above medium, which suggested that they are acceptable. The average scores for boys are between 1.42 and 1.50 and those for girls are between 2.22 and 2.28. The gap of the average is close to 1. This phenomenon matches the results of Hwang et al. (2013b) who found that the cognitive load and competition anxiety of females are higher than those of males.

5.2. Impact on learning performance

To explore whether the learning performances of the different genders are significantly different, we conducted ANCOVA analysis. In the analysis, the pre-test was regarded as a co-variate, the post-test as a dependent variable, and gender as an independent variable. First, the interaction effects of the independent variable and the co-variate were observed. The results showed that the between-group and the pre-test of the ANCOVA analysis is not significant ($F = 3.422$, $p = .078 > .05$). This means that the independent variable and the co-variate do not interactively affect the result, so the analysis could be continued. For the two genders, the ANCOVA analysis of excluding the impact of the pre-test is shown in Table 4.

Table 4. ANCOVA results of the two genders after excluding the impact of the pre-test

Groups	Number	Average	*SD*	Adjusted average	*F*	*p*
Boys	12	9.08	2.353	8.997	2.579	0.123
Girls	12	7.42	2.968	7.503		

From Table 4, it is seen that the adjusted average scores of the post-tests are 8.997 and 7.503 for the boys and girls respectively. It is seen that after excluding the impact of the pre-test, the two genders did not reach significant difference ($F = 2.579$, $p = .123 > .05$). Although the difference is not significant, boys' learning performance is higher than that of girls.

5.3. The co-impact of normalized gain score and mental effort score

Paas and van Merriënboer (1993) proposed learning efficiency, which is measured by the value of the normalized performance score subtracting the mental effort score. In this study, we used the gain score to represent the performance score. The gain score is the score of the post-test subtracting the score of the pre-test. To normalize the gain score, we used the linear equation of A*X+B=Y, where X represents the gain score and Y represents the normalized gain score, as the normalized formula. Then we mapped the maximum normalized gain score to 5 (i.e., maximum score of the ME questionnaire) and the minimum normalized gain score to 1 (i.e., minimum score of the ME questionnaire). Sequentially, two equations were obtained. We solved the two equations and found A and B. In our case, for the 24 valid samples, the maximum value of the gain scores is 6 and the minimum value is (-3). The two equations are as follows:

$$6*A+B=5$$
$$(-3)*A+B=1$$

By solving the two equations, the results are that A equals (4/9) and B equals (7/3). Finally, we normalized the gain score with the formula of (4/9)*X+(7/3)=Y. According to this formula and the learning efficiency equation of LE=Y-ME, we could calculate the normalized gain score for each child and then the learning efficiency could be calculated. Sequentially, the independent samples *t* test of the learning efficiency of the boys and girls could be analyzed. The results are shown in Table 5.

Table 5. The independent samples *t* test for analyzing learning efficiency

Sex	Number	Average	*SD*	*t*	*Cohen's d* (Effect size)
Boys	12	2.067	1.032	2.915^{**}	0.842
Girls	12	0.575	1.441		

Note. $^{**}p < .01$.

From the table, it can be seen that the boys' learning efficiency was significantly higher than that of girls ($t = 2.915$, $p < .01$, $d = 0.842$). Accordingly, the effect size of the aforementioned results is large, which suggested that they are acceptable. This means that the instructional design benefits boys more than girls.

6. Discussion

From the above experimental results, it can be seen that the cognitive load of boys is significantly lower than that of girls. Although the difference in learning performance is not significant, that of boys is higher than that of girls. After further observation, the learning efficiency obtained after subtracting the mental effort score from the normalized gain score, the average score of boys is significantly higher than that of girls. These results represent that the use of educational robots to learn idioms for elementary school students is more beneficial for boys than for girls. By observing the data, there were four students whose gain scores were negative. These four students were all girls. We therefore infer that for some girls, ZILS may cause disturbances to their learning. As the gain scores are all positive for boys, we infer that boys are generally more interested in emerging technologies and games. This result is similar to the finding of Hwang et al. (2013b) who found that for elementary school students, girls have more cognitive load and competition anxiety than boys.

The reasons behind these phenomena can be explained by referring to the literature. For example, Pauls, Petermann, and Lepach (2013) found that male students are more interested in visual and spatial aspects, while female students are more interested in the hearing aspect. Herlitz and Rehnman (2008) obtained similar results. The feelings of the boys when using ZILS may be related to the cute shape of the robot, the pictures in the system, and the placement of the robot in the classroom, which may bring changes to the classroom space. These are attractive to boys. Emerging technologies and games may cause girls learning interference. Webley (1981) also pointed out that boys explore new environments more frequently than girls. The author's argument also explains our results: boys may be more interested in exploring changes in the environment. Besides, Master et al. (2017) indicated that young girls had less interest and self-efficacy in technology compared with boys in elementary school. Ring (1991) also indicated that male students had greater self-confidence than female students in their ability to use courseware (and hence computers) as an effective learning tool. Dweck, Davidson, Nelson, and Enna (1978) indicated that females showed greater evidence of non-adaptive attributions and therefore were more adversely affected by failure. Therefore, the integration of emerging technologies and games into teaching will be of more benefit to boys.

With more in-depth discussion, our system has two major characteristics of game pictures and robotics technology. It brings changes in the look of the classroom space, with a pleasing appearance, which can arouse the interest of boys with strong sensory abilities such as vision and space. Girls with strong hearing ability may prefer to study quietly and have a greater cognitive load on general scientific and technological teaching activities. However, if science and technology teaching activities can be designed to have a strong spiritual level and have the connotation of deep learning in the process, it is likely to help girls reduce their cognitive load when using technology for learning. Therefore, although this research developed one unit only, in the future, two more units can be designed without duplication of idioms, and the difficulty of guessing the idioms in the three units can increase in order. For example, the first unit uses "character" as the reminder object to test children's memory of words. The second unit could use "idiom story" as the reminder theme to test students' memory and understanding of the storyline, while the third unit could use "idiom sentence-making" as a reminder theme to test students' understanding of the application of sentence-making. The game could be designed to interact with the robot in a complex manner, for example, when the idiom story is a reminder, not only could the picture be used, but the robot's NLP function could also be designed to interact with the students in the storyline picture. In the question/answer interaction, students would not only be guessing idioms, but also people and things in the story. Such learning content would not only improve the quality of the game elements, but would also make full use of the NLP function of the modern AI of robots. It may be able to help girls improve their learning effectiveness and reduce their cognitive load, thereby enhancing their learning efficiency.

Based on such future improvements, our system will be able to provide more real-time interaction and contextual learning to increase elementary school students' interest in learning Chinese idioms when compared with the past studies which have explored how to increase elementary school students' interest in learning Chinese idioms (Ku, Huang, & Hus, 2015; Wong, Chin, & Tan, 2010). The impacts of gender differences on other indicators of cognitive load, such as germane cognitive load (Lange & Costley, 2019) can be further verified.

7. Conclusions and future works

This study explored the topic of "gender differences in cognitive load on applying game-based learning by intelligent robots". To achieve this aim, this study developed a game-based learning system based on Zenbo robots, which uses the sound and the cute image of the Zenbo robot to attract elementary school students' interest in the learning content. During the learning process, the first module in ZILS is "idiom learning." It uses a homepage showing animal images, and lets Zenbo read out the screen's text to impress the students with the idiom story. The second module is "reviewing with games." It uses creative pictures from the game as a reminder for students to guess the idioms. This design can refresh the students' memory. We also explored the impacts of gender differences on the related aspects of CLT for elementary students in using the system. The results show that this system is more beneficial for boys. Their cognitive load is significantly lower. Their learning performance is also better, although it does not reach significance. Furthermore, the learning efficiency for them is significantly higher. From the literature, we also found that males and females have different learning interests. Males are mainly interested in the visual and spatial aspects while females are mainly interested in the hearing aspect. These may be the reasons for the differences in the results.

In the future, if the system can be improved by increasing the game elements and the interaction with the robot's NLP function, it will be able to benefit girls and may have deeper findings. In addition, when using emerging technologies for E-learning, boys and girls have different physical, psychological and learning characteristics (Liao, Zhen, & Hwang, 2018). If more information of the learning process can be recorded and the impact on students' behavior can be explored, it should be possible to sort out more helpful rules and can serve as optimization criteria for the development of learning systems. Therefore, gender differences in behavior analysis in the environments of educational robots will be an important research issue that is worth exploring in the future, in order to optimize the system and improve students' learning performance.

Acknowledgement

This study is supported by the Ministry of Science and Technology of Taiwan under contract number MOST 107-2511-H-275-001 and MOST 108-2511-H-224-006-MY3.

References

Barak, M., & Zadok, Y. (2009). Robotics projects and learning concepts in science, technology and problem solving. *International Journal of Technology and Design Education, 19*(3), 289–307.

Bevilacqua, A. (2017). Commentary: Should gender differences be included in the evolutionary upgrade to cognitive load theory? *Educational Psychology Review, 29*(1), 189–194.

Cai, J. S., Yan, R. C., Yang, J. J., & Wang, R. Y. (2012, May-June). *A Study of individual difference on learners' acceptance toward game-based instructional design of 3D animation*. Paper presented at the 16th Global Chinese Conference on Computers in Education, Pingtung, Taiwan.

Chen, B., Hwang, G. H., Wang, S. H., & Peng, J. H. (2018, March). *The Development of an intelligent robot navigation system - A case study of Du-Xing elementary school located in north district of Taichung of Taiwan*. Paper presented at the 3rd International Conference on Digital Learning Strategies and Applications (DLSA), Hokkaido, Japan.

Chen, X., Xie, H., & Hwang, G. J. (2020). A Multi-perspective study on artificial intelligence in education: Grants, conferences, journals, software tools, institutions, and researchers. *Computers & Education: Artificial Intelligence, 1*, 100005. doi:10.1016/j.caeai.2020.100005

Christophel, E, & Schnotz, W. (2017). Gender-specific covariations between competencies, interest and effort during science learning in virtual environments. *Frontiers in Psychology, 8*, 1681. foi:10.3389/fpsyg.2017.01681

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates, Publishers.

Crossley, S. A., Allen, L. K., Snow, E. L., & McNamara, D. S. (2016). Incorporating learning characteristics into automatic essay scoring models: What individual differences and linguistic features tell us about writing quality. *Journal of Educational Data Mining, 8*(2), 1-19.

Dweck, C. S., Davidson, W., Nelson, S., & Enna, B. (1978). Sex differences in learned helplessness: II. The Contingencies of evaluative feedback in the classroom and III. An experimental analysis. *Developmental Psychology, 14*(3), 268-276.

Gates, B. (2007). *Everybody has a robot (in Chinese)* [Scientific People Magazine]. Retrieved from http://sa.ylib.com/MagArticle.aspx?Unit=featurearticles&id=966

Hart, S. A. (2016). Precision education initiative: Moving toward personalized education. *Mind, Brain, and Education, 10*(4), 209-211.

Herlitz, A., & Rehnman, J. (2008). Sex differences in episodic memory. *Current Directions in Psychological Science, 17*(1), 52–56.

Hsiao, H. S., Lin, Y. W., Lin, K. Y., Lin, C. Y., Chen, J. H. & Chen, J. C. (2019). Using robot-based practices to develop an activity that incorporated the 6E model to improve elementary school students' learning performances. *Interactive Learning Environments*. doi:10.1080/10494820.2019.1636090.

Hsiao, J. M., & Huang, Y. C. (2012). An Exploratory study of entrepreneurship in applying LEGO MINDSTORMS® robot on science and creativity education (in Chinese). *Journal of Far East University, 29*(3), 375-386.

Huang, Y. M., Shadiev, R., & Hwang, W. Y. (2016). Investigating the effectiveness of speech-to-text recognition applications on learning performance and cognitive load. *Computers & Education, 101*, 15-28.

Hwang, G. J., Sung, H. Y., Chang, S. C., & Huang, X. C. (2020). A Fuzzy expert system-based adaptive learning approach to improving students' learning performances by considering affective and cognitive factors. *Computers & Education: Artificial Intelligence, 1*, 100003. doi:10.1016/j.caeai.2020.100003

Hwang, G. J., Xie, H., Wah, B. W., & Gašević, D. (2020). Vision, challenges, roles and research issues of artificial intelligence in education. *Computers & Education: Artificial Intelligence, 1*, 100001. doi:10.1016/j.caeai.2020.100001

Hwang, G. J., Yang, L. H., & Wang, S. Y. (2013a). A Concept map-embedded educational computer game for improving students' learning performance in natural science courses. *Computers & Education, 69*, 121-130.

Hwang, M. Y., Hong, J. C., Cheng, H. Y., Peng, Y. C., & Wu, N. C. (2013b). Gender differences in cognitive load and competition anxiety affect 6th grade students' attitude toward playing and intention to play at a sequential or synchronous game. *Computers & Education, 60*(1), 254-263.

Javora, O., Hannemann, T., Stárková, T., Volná, K. & Brom, C. (2019). Children like it more but don't learn more: Effects of esthetic visual design in educational games. *British Journal of Educational Technology, 50*(4), 1942-1960.

Kovecses, Z., & Szabo, P. (1996). Idioms: A View from cognitive semantics. *Applied Linguistics, 17*(3), 326–355.

Krell, M. (2017). Evaluating an instrument to measure mental load and mental effort considering different sources of validity evidence. *Cogent Education, 4*(1), 1280256. doi:10.1080/2331186X.2017.1280256

Ku, D. T., Huang, Y. H., & Hus, S. C. (2015). The Effects of GBL and learning styles on Chinese idiom by using TUI device. *Journal of Computer Assisted Learning, 31*(6), 505-515.

Kubilinskiene, S., Zilinskiene, I., Dagiene, V., & Sinkevicius, V. (2017). Applying robotics in school education: A Systematic review. *Baltic Journal of Modern Computing, 5*(1), 50-69.

Kucuk, S. & Sisman, B. (2017). Behavioral patterns of elementary students and teachers in one-to-one robotics instruction. *Computers & Education, 111*, 31-43.

Lange, C. & Costley, J. (2019). How sequencing and fading affects the relationship between intrinsic and germane cognitive loads. *Distance Education, 40*(2), 243-261.

Liao, C. W., Chen, C. H. & Shih, S. J. (2019). The Interactivity of video and collaboration for learning achievement, intrinsic motivation, cognitive load, and behaviour patterns in a digital game-based learning environment. *Computers & Education, 133*, 43-55.

Liao, Y. W., Zhen, X. J., & Hwang, G. H. (2018, November). *Exploring students' learning outcome and gender differences in a digital video clip course*. Paper presented at the 26th International Conference on Computers in Education, Manila, Philippines.

Liddy, E.D. (2001). Natural language processing. In *Encyclopedia of Library and Information Science* (2nd ed., pp. 1-15). New York, NY: Marcel Decker, Inc.

Liu, E. Z. F., & Lin, C. H. (2009). Developing evaluative indicators for educational computer games. *British Journal of Educational Technology, 40*(1), 174-178.

Luckin, R., Holmes, W., Griffiths, M., & Forcier, L. B. (2016). *Intelligence unleashed: An Argument for AI in education*. London, UK: UCL Knowledge Lab.

Master, A., Cheryan, S., Moscatelli, A., & Meltzoff, A. N. (2017). Programming experience promotes higher STEM motivation among first-grade girls. *Journal of Experimental Child Psychology, 160*, 92-106.

Mayer, R. E. (2009). *Multimedia learning, second edition*. New York, NY: Cambridge University Press.

Mutlu-Bayraktar, D., Cosgun, V., & Altan, T. (2019). Cognitive load in multimedia learning environments: A Systematic review. *Computers & Education, 141*, 103618.

Ng'ambi, D. (2013). Effective and ineffective uses of emerging technologies: Towards a transformative pedagogical model. *British Journal of Educational Technology, 44*(4), 652–661.

Paas, F. G. W. C., & van Merriënboer, J. J. G. (1993). The Efficiency of instructional conditions: An Approach to combine mental effort and performance measures. *Human Factors, 35*, 737–743.

Paas, F. G. W. C., & van Merrienboer, J. J. G. (1994). Instructional control of cognitive load in the training of complex cognitive tasks. *Educational Psychology Review, 6*(4), 351–371.

Paas, F. G. W. C., Tuovinen, J. E., Tabbers, H., & van Gerven, P. (2003). Cognitive load measurement as a means to advance cognitive load theory. *Educational Psychologist, 38*(1), 63–71.

Paas, F. G. W. C., van Gog, T., & Sweller, J. (2010). Cognitive load theory: New conceptualizations, specifications, and integrated research perspectives. *Educational Psychology Review, 22*(2), 115–121.

Papadopoulos, I., Lazzarino, R., Miah, S. & Weaver, T. (2020). A Systematic review of the literature regarding socially assistive robots in pre-tertiary education. *Computers & Education, 155*, 103924.

Pauls, F., Petermann, F., & Lepach, A. C. (2013). Gender differences in episodic memory and visual working memory including the effects of age. *Memory, 21*(7), 857–874.

Piaget, J. (1964). *Part I: Cognitive development in children: Piaget development and learning*. New York, NY: Wiley Periodicals, Wiley Company.

Resnick, M., & Ocko, S. (1991). LEGO/Logo: Learning through and about design. In I. Harel & S. Papert (Eds.), *Constructionism* (pp. 141-158), Norwood, NJ: Ablex Publishing.

Ring, G. (1991). Student reactions to courseware: Gender differences. *British Journal of Educational Technology, 22*(3), 210-215.

Rusk, N., Resnick, M., Berg, R., & Pezalla-Granlund, M. (2008). New pathways into robotics: Strategies for broadening participation. *Journal of Science Education and Technology, 17*(1), 59–69.

Smith, G. G., Haworth, R., & Žitnik, S. (2020). Computer science meets education: Natural language processing for automatic grading of open-ended questions in eBooks. *Journal of Educational Computing Research, 58*(7), 1227-1255.

Sweller, J. (1998). Cognitive load during problem solving: Effects on learning. *Cognitive Science, 12*(2), 257-285.

Sweller, J. (2005). Implications of cognitive load theory for multimedia learning. In R. E. Mayer (Ed.), *The Cambridge Handbook of Multimedia Learning (pp.19-30)*, New York, NY: Cambridge University Press.

Varney, M. W., Janoudi, A., Aslam, D. M., & Graham, D. (2012). Building young engineers: TASEM for third graders in woodcreek magnet elementary school. *IEEE Transactions on Education, 55*(1), 78–82.

Webley, P. (1981). Sex differences in home range and cognitive maps in eight-year-old children. *Journal of Environmental Psychology, 1*(4), 293–302.

Westlund, J. M. K., Jeong, S., Park, H. W., Ronfard, S., Adhikari, A., Harris, P. L., DeSteno, D. & Breazeal, C. L. (2017). Flat vs. expressive storytelling: Young children's learning and retention of a social robot's narrative. *Frontiers Human Neuroscience, 11*, 295.

Wong, L. H., Chin, C. K., & Tan, C. L. (2010). Students' personal and social meaning making in a Chinese idiom mobile learning environment. *Educational Technology & Society, 13*(4), 15-26.

Wong, M., Castro-Alonso, J. C., Ayres, P., & Paas, F. (2015). Gender effects when learning manipulative tasks from instructional animations and static presentations. *Educational Technology & Society, 18*(4), 37–52.

Wu, W. C. V., Wang, R. J., & Chen, N. S. (2015). Instructional design using an in-house built teaching assistant robot to enhance elementary school English-as-a-foreign-language learning. *Interactive Learning Environments, 23*(6), 696-714.

Yu, G. Z. (1994). *From Xu Xiake to Van Gogh (in Chinese)*. Taipei City, Taiwan: Chiu Ko Publishing Co.

Zhong, B., Zheng, J. & Zhan, Z. (2020). An Exploration of combining virtual and physical robots in robotics education. *Interactive Learning Environments*. doi:10.1080/10494820.2020.1786409