iAbstract: Game-driven Keyword Auction and Summarization for Academic Reading

Hercy N. H. Cheng[1], Calvin C. Y. Liao[2*] and Wan-Chen Chang[3]

[1]National Engineering Research Center for E-Learning, Central China Normal University, China // [2]College of Nursing, National Taipei University of Nursing and Health Sciences, Taiwan, R.O.C // [3]Department of Human Development and Family Studies, National Taiwan Normal University, Taiwan, R.O.C // hercycheng.tw@gmail.com // CalvinCYLiao@gmail.com // altheawcc@ntnu.edu.tw

[*]Corresponding author

ABSTRACT

Graduate students usually lack sufficient ability to read academic papers effectively and efficiently. For facilitating their reading comprehension, this study adopts summarization as a main reading strategy under the scaffold of keyword evaluation. Furthermore, this study attempts to transform summarization into a group-based educational game by incorporating keyword auction mechanisms. The research purpose is to investigate how first-year graduate students react to such an educational game when they are required to summarize a section of a real journal paper. The results indicate that they improve the completeness of the summaries without decreasing their conciseness after the game. Besides, the analysis of their prior summaries suggests that graduate students may lack the ability of structure analysis, idea integration and argumentation. Additionally, the results also show that there are positive relationships among the students' evaluation, experts' evaluation, auction behaviors and frequencies in summaries of the keywords, suggesting that students' keyword evaluation may change their decision-making behaviors and the quality of their final summaries.

Keywords

Academic reading, Summarization, Game-based learning, Auction mechanism

Introduction

While undergraduate education usually focuses on teaching well-organized knowledge, graduate education allows students to explore emerging knowledge. Because graduate students have to conduct real and innovative research, they should learn how to do research in an authentic academic context so that they could develop their academic reading and writing abilities. These abilities involve tacit knowledge, which is usually hard to learn by lectures. This is why graduate education usually adopts apprenticeship, where students can learn from their advisors and academic papers. Furthermore, they read literature, discuss with their advisors, conduct a series of researches, and write academic papers. Through the process, they find their own purposes and meanings. The authenticity is necessary for graduate students to absorb the knowledge and know-how of doing research (Herrington, Reeves, & Oliver, 2014).

Brown, Collins, and Duguid (1989) described the authenticity as meaningful, purposeful, but ordinary practices of the culture. In general, the researchers of educational technology realized the authenticity in two ways: establishing authentic environments and providing authentic problems (see Herrington & Oliver, 2000). In authentic learning environments, students are allowed to interact with real or simulated people and artifacts. For doing so, researchers adopted the technology of virtual or augmented reality so that students may learn in real situations. For example, Huang, Yang, Chiang, and Su (2016) allowed students to use their own mobile phones to learn English vocabulary in outdoor learning sites. On the other hand, authentic problems require students to enquiry and solve contextual learning tasks, so that they may apply and reflect the knowledge in real life (Maina, 2004). For example, Hwang, Chiu, and Chen (2015) designed a contextual educational game, which could situate students in a series of inquiry quests in social studies. In a sense, game provided a reasonable context so that students may feel more engaged in solving complicated learning tasks.

Although authentic problems usually preserve the complexity of the real-life setting, these problems are relevant to students. In the process of problem solving, the students may practice their expertise and share with others. Like in graduate education, for doing appropriate research and writing academic papers, a fundamental capability is to read sufficient related papers. Furthermore, graduate students have to locate, organize, and explain important knowledge in their own way and for their own research purposes. However, several thousand Chinese students in a university reported their difficulties of getting the meaning of the article and determining the main ideas (Evan & Green, 2007), because novice academic readers usually lack prior knowledge to distinguish the contributions of the authors from previous works. They also do not have a specific goal to locate relevant content in the beginning. In this vein, grasping the gist and identify useful information become the first challenge of the first-year graduate students.

Among various reading strategies, summary writing may well facilitate students' reading comprehension by synthesizing important ideas in papers (Carroll, 2008), while keywords evaluation may help students grasp the main ideas more easily (Chou, 2012). Although researchers have been developed many effective systems for learning how to read better, the effects seem decreased over time, especially for low-ability students (Magliano et al., 2005). After using these strategies for a while, the students may easily feel unmotivated. For these reasons, this study aims to integrate summarization with keyword evaluation in order to design a group-based game, iAbstract, for graduate students to summarize academic papers. In a sense, this study gamifies the process of academic summarization with several techniques of text analysis. Furthermore, this study adopts Chinese word segmentation technique to extract and weight keywords and latent semantic analysis to calculate the scores of summaries for supporting the gameplay. However, the educational technology researchers' endeavor to gamify reading strategies is still in an early stage, because the reading and writing abilities could not be easily evaluated if researchers would like to avoid explicit assessment in games. Recently, as text analysis continues to a mature technology, the research on game-based reading and writing become more feasible. At this stage, it is expected that this study may enrich our understanding of the relationship between students' game behaviors and learning outcomes. Besides, this study also concerns how graduate students compose their academic summaries in such a game-based learning environment.

Related works

Academic reading and summarization

Recent studies on academic reading focus on students' learning experience more than learning performance (e.g., Liu, Chang, Yang, & Sun, 2011). Furthermore, some studies adopted reading-writing connection activities to improve students' academic reading performance. For example, Shih (1992) classified academic reading lessons as three main phases: pre-reading, guided reading and post-reading exercises. Among the three phases, post-reading exercises, such as note taking and summarizing, are useful strategies for reducing details and organizing important ideas in papers. Baker and Boonkit (2004) found that the frequently used strategies for academic reading and writing were general problem-solving strategies, such as reading for the main idea of passages, connecting prior knowledge and new one during writing, etc. Besides, Kim and Kim (2017) found an integrated reading-to-write task could help graduate students' capabilities of getting main ideas from the academic articles. In other words, writing activities after academic reading could help students' learning performance.

Summarization, or summary writing, is a post-reading activity. Previous studies have shown that students who participated in summary writing have better reading comprehension than those who did not (Carroll, 2008). This is due to the fact that summarization not only requires students to read the article carefully, but also need them to understand its content more deeply. Furthermore, for writing summaries, students have to identify the important elements of the article and integrate the past knowledge with the new one (Wade-Stein & Kintsch, 2004). In other words, in order to write better summarization, students should comprehend the article first. Besides, summarization may also guide students to focus on the important ideas of the article and to build the relationships among them. By doing so, when summarizing an article, students can make better deletion, retention and modification of the content. In this process, students can reorganize the ideas and then gain better comprehension.

In general, the theories of summarization suggested that students should carry out the process of material deletion, word composition, and topic sentence invention (McNeil & Donant, 1982). In material deletion, students have to evaluate which information is unnecessary and redundant for the gist of articles. In other words, they determine the importance of the materials in articles. In word composition, students are encouraged to select and adopt a series of short words to replace the remained sentences. For example, Chou (2012) adopted keywords and questions. More specifically, after reading an article, students had to write three keywords that they thought important, and to answer 5W1H questions (i.e., what, why, where, when, who, and how) for writing summaries. By doing so, students can highlight what they should pay attention. Later, in topic sentence invention, students are required to write topic sentences to contextualize those words for reorganizing the gist of the articles.

Computer-based feedback for summarization

Because summarization is so important for reading comprehension, many researchers have designed various learning activities and scaffolds, so that students can write high-quality summaries (Konuk, Ören, Benzer, &

Sefer, 2016). Although it is critical for students to practice summarization, grading students' summarization or providing or feedback are still time-consuming and laborious tasks for teachers. Therefore, many studies adopted a variety of computer-based techniques to overcome this problem (Kintsch et al., 2000; Wade-Stein and Kintsch, 2004; Landauer, Lochbaum, & Dooley, 2009; Sung, Liao, Chang, Chen, & Chang, 2016). For example, He, Hui, and Quan (2009) reviewed several advanced techniques for supporting computer-based feedback of summarization, such as latent semantic analysis (LSA; Kintsch et al., 2000). The LSA has been adopted to design computer-based feedbacks, so that teachers may carry out the grading of summaries more efficiently.

For example, Kintsch et al. (2000) designed a first prototype of a computer-based feedback system, called *State the Essence*. In the system, they adopted the LSA technique to calculate the relationships between students' summaries and source texts. Besides, the system also provided students with feedback for improving summaries, such as spelling, length, and overall content coverage. Wade-Stein and Kintsch (2004) then developed a *Summary Street* system to assess students' summaries and provide feedback. Furthermore, they indicated that the students with feedback might spend more time on writing their summaries. Besides, they also found that the feedback demonstrated a lasting effect on students' ability to summarize articles. In addition to providing feedback, researchers also have begun to integrate different scaffoldings in the past decade. For example, Landauer, Lochbaum and Dooley (2009) further developed *WriteToLearn*, a computer-supported writing system for reading comprehension. Moreover, the system provided students with tutorial feedbacks according to their traits of writing. In recent year, Sung et al. (2016) integrated concept maps to provide students with semantic and conceptual scaffolds. Their finding showed that the proposed system could improve the students' summary writing skills.

Digital game-based reading activities

For engaging students in learning strategies for reading, many research designed various virtual characters. For example, McNamara, Levinstein, and Boonthum (2004) designed an *iSTART* system for training students' reading strategies. In the system, more specifically, three virtual characters, including an intelligent tutor and two virtual students, demonstrated reading strategies for learners. These learners may learn how to paraphrase, elaborate and explain the meaning of selected articles (McNamara, 2009). Besides, Park and Kim (2015) developed a *Virtual Tutee System* to enhance college students' academic reading engagement. In the system, after reading assigned articles, the college students were required to teach virtual tutees by providing lecture notes and answering the tutees' questions.

The concept of virtual characters stimulates the research of game-based learning. Furthermore, some researchers started to incorporate game-based interaction among students. For example, Dempsey, Jackson, and McNamara (2010) integrated game characteristics into the iSTART system and developed *MiBoard*, a multiplayer online board game. In this game, students played the role of an explainer in turn. The explainer was required to use randomly selected strategies to explain a sentence from an article, while the other players guessed which one strategy the explainer used. Given that there was not much research on digital game-based reading, the MiBoard system inspired our design. Furthermore, game mechanisms are used to provide students with clear and positive goals for writing better summaries, while group-based activities may facilitate interactions among players.

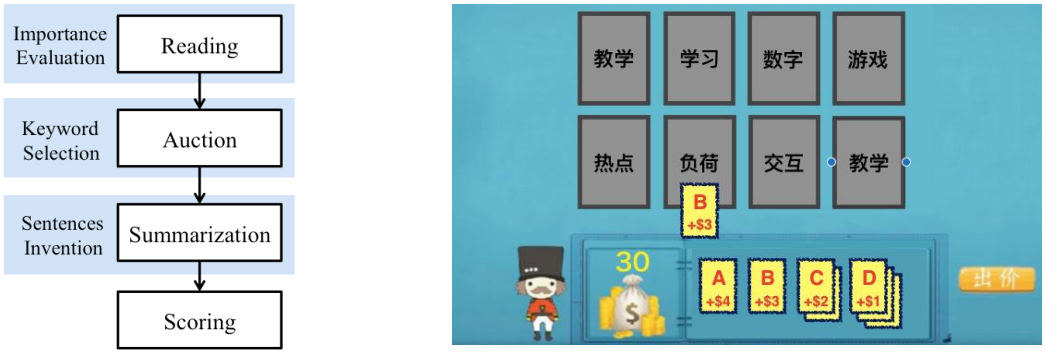
Game design

Design framework

As McNeil and Donant (1982) suggested, a tutorial approach to summarization includes material deletion, word composition, and topic sentence invention. According to this approach, this study adopts a design framework, illustrated in Figure 1(a). In this framework, for writing summaries, students are required to evaluate the importance of information in articles, select appropriate keywords, and composing topic sentences. Following this framework, the study designed an educational game by integrating summarization and keyword auction to support students to participate in summary writing with pleasure. Furthermore, the summarization game requires every student to read an article, acquire important keywords, and then use the keywords to compose a summary.

An auction game mechanism is applied to the process of keyword selection. More specifically, students attempt to offer higher bids for high-valued keywords than the others. Meanwhile, they also have to control their costs to avoid offering too much bids. For doing so, they have to mentally evaluate the importance of keywords and

make correct decisions on bidding. In a sense, the auction mechanism is to compare students' evaluation of keywords with others.



(a) The design framework (b) The interface of the auction

Figure 1. Game design

The game consists of four phases: reading, auction, summarization, and scoring.

(1) Reading: Students use their own personal devices to read an academic article within a limited time. Meanwhile, the students are also required to evaluate selected keywords. For quantifying the importance of the keywords, this study adopted the Chinese word segmentation technique of Nature Language Processing and Information Retrieval Platform (NLPIR) of Chinese Academy of Science to produce a set of keywords with weights, which were calculated according to their frequencies and the similarities to the semantic space of the article.

(2) Auction: Students are required to bid a reasonable price for important keywords. Furthermore, they plan auction on their personal devices, and then perform the action face to face. The auction mechanism will be introduced in detail in the next section. In the end of the auction, their devices show the selected keywords with different scores, which are assigned secretly depending on the weights of keywords in the article.

(3) Summarization: Students are required to write summaries with keywords. Furthermore, they have to use the purchased keywords, but they are not allowed to use the keywords they do not buy. This rule is designed to make students select keywords more seriously in the previous stage. If they did not follow the rule, they would be asked to re-write the summaries.

(4) Scoring: Finally, the game score is consisted of keywords importance, costs, and summary qualities. The keyword importance is equal to the scores of the purchased keywords (ranging 2 to 5 points per keyword). The costs are evaluated by students' remaining money (1 points for every three remaining coins). The summary qualities are automatically calculated by the technique of latent semantic analysis for the gameplay. More specifically, the similarity of every sentence in their summaries to the article was calculated and transferred into 15 points at most. Therefore, for wining this game, students have to write better summaries, purchase higher-value keywords and pay lower prices. Finally, the player with the highest score wins the game.

Auction mechanism

In the auction phase, the students are given 30 game coins and start to plan their auction. The auction planning allows students to consider all keywords in advance at the same time, instead of bidding one keyword at a time. Such design not only requires the students to have a whole picture, but also facilitates them to compare the importance of these keywords, lessening the influence of other factors, such as the cost. Furthermore, they are given a set of bidding cards, including one A class card, two B class cards, two C class cards, and three D class cards, as shown in Figure 1(b). The A card allows the students to acquire the keyword they wanted without any competition, but this card may increase its price by four coins. The B cards, which may increase the prices of the keywords by three coins, allow the students to acquire the keywords only when the other players do not use any A cards to bid. Similarly, the C bidding cards may increase the prices by two coins and allows them to acquire the keywords only when the others do not use any A or B bidding cards. The D cards are used to specify that the students do not want the keywords, but the cards may still increase the prices by one coin. Such design makes sure that students can acquire at least one keyword, avoiding the problem of failed auction.

After all students plan their auction secretly, the system calculates the prices of the keywords according to all bidding cards. After the prices are displayed, every highest bidder has to decide whether he/she wants to buy the keywords at the price. If he/she wants the keyword, then he/she has to pay the price. If not, the bidders with next class cards have the chance to decide whether to buy it. Furthermore, although students offer prices at the same time, the prices are varied and aggregated from all students' evaluation. In other words, if all students wanted the same keyword, the price becomes higher. In a sense, the prices may reflect students' evaluation of importance, which may allow them to compare their own evaluation with each other.

As mentioned earlier, the keywords are assigned with different scores, ranging from 2 points to 5 points. For preventing them from the influence of the scores, the scores are revealed in the last phase. Besides, in order to facilitate them to buy as more important keywords as possible, those coins will be devalued if they are not spent. Students have to evaluate whether the prices are worth enough. In an ideal case, students, who are able to exchange coins to keywords with equivalent values, have to use these keywords to compose better writings for winning. Otherwise, if any other players underestimate a keyword, they may have a chance to score more. Alternatively, if other players overestimate it, they may decide not to buy it and save the money to buy other keywords.

Method

Research questions

This study attempts to investigate how students compose their summaries and how they make evaluations in such an educational game. Furthermore, this study examines the relationships among students' comprehension of the articles and auction behaviours as well as how these factors influence their summary writing. More specifically, there are two research questions: (1) How does the game change students' summaries? (2) What are the relationships among students' evaluation of keywords, summaries, and auction behaviours?

Reading material

Article selection

This study selected the *Introduction* section of a Chinese journal paper, entitled "*The study of flow experience in digital game-based learning*" (Chen, Yang, & Wen, 2016), as the learning material. This article was published at a Chinese Social Science Citation Indexed journal "Digital Education Research." The article aimed to examine the influences of students' academic ability and social interaction tendency on their flow experience in digital game-based learning. Before the study, the participants were asked about their knowledge of flow experience. The main ideas of this article are listed in Table 1. Each main idea should have an argument with explanation.

Table 1. The main ideas of summaries

Categories	Main Ideas
Research motivation	Argument: The reason for studying flow experience Explanation: Definition of flow experience
Hypothesis of variable 1	Argument: The hypothesis of academic ability Explanation: The perspectives of flow theory
Hypothesis of variable 2	Argument: The hypothesis of social tendency Explanation: The perspectives of motivation theory
Hypothesis of interaction	Argument: The hypothesis of interactions between the two variables Explanation: The perspectives of cognitive load
Research objective	Argument: The objective for studying the influence of the two variables and their interaction on flow experiences Explanation: The absence of previous studies

Keywords extraction

This study extracted and selected eight keywords from the article for the game. More specifically, this study adopted the NLPIR Platform of Chinese Academy of Science to produce the top 50 important keywords of the article. As sown in Table 2, the researchers selected two keywords ("games" and "digital") from top 10 keywords

and assigned them as 5 game points. Similarly, the keywords with 4 and 3 points were selected from top 11-20 keywords and top 21-50 keywords, respectively. Finally, the researchers selected additional two keywords that were not included in top 50 keywords and assigned them as 2 points.

Table 2. Keywords selection for the game

Rank of weights	Selected keywords	Meaning in the article	Game scores
Top 1-10	Games	Research context: digital games, game-based learning (environments)	5
	Digital	Research context: digital games, digital game-based learning	5
Top 11-20	Learning	Research context: learning in games	4
	Load	Backup arguments for research purposes: the perspectives of cognitive load	4
Top 21-50	Teaching	Research context: Instructional activities, applying game-based learning to instruction	3
	Interaction	Research purposes: statistical interactions between two independent variables	3
Not in top 50	Internal	Descriptions of cognitive load	2
	Hot issues	Descriptions of game-based learning	2

Procedure

The study involved 32 first-year graduate students (8 males and 24 females). Each session of the experiment recruited four students. Before a game, the students were required to answer questions about their prior knowledge of the topic. There were 29 (90.62%) students who did not correctly explain the knowledge, while only 3 students who were fully or partially correct about it. As a coordinator, a researcher then introduced the game rules and started the activity. In the reading phase, the students had to read the article in 10 minutes. Besides, they also had to sort the selected eight keywords according to their perceived importance to the article and write down prior summaries on their devices. The prior summaries were used to help them preliminarily draft the main ideas of this paper. Meanwhile, the researcher could collect their initial thoughts about this paper. In the auction phase, they planned their auction secretly and then decided whether to buy keywords. Highest bidders paid the prices and used the keywords to compose final summaries, which must included five sentences. After the scoring phase, finally, they were also required to individually explain their behaviors in the game.

Data collection and analysis

During the game, students' summaries, evaluation of keywords, and behaviors of auction are all collected. Students' summaries were denoted by the format like "S11-final." The "S11" represented the identification number of one student, while the "final" represented that it was a final summaries rather than a prior one. Besides, two independent raters evaluated the quality of students' prior and final summaries according to Table 3. As mentioned earlier, article was consisted of five main ideas, starting from research motivation, followed by three hypotheses and concluded with research objective. As shown in Table 3, a complete main idea should include a concrete argument with detailed explanation. Kappa statistic was adopted as a measure of agreement between the raters. The Kappa values were .855, suggesting a consistent agreement.

Table 3. The criteria of each main idea in summaries

Quality	Definition	Scores
Complete	A concrete argument with detailed explanation	4
Flawed	Only an argument without detailed explanation	3
Incomplete	Only explanation without a concrete argument	2
Unrelated	Statements unrelated to the main idea	1
None or Wrong	Not mentioned any statements or wrong statements	0

Besides, this study also adopted terse values to evaluate the terseness of students' writing, such as note taking or summary writing (Chang, & Ku, 2015). The terse values could be regarded as an informational density, indicating whether the students could sufficiently elaborate the main ideas by composing brief words. The terse values were calculated by the number of main ideas divided by the number of words (i.e., Chinese characters, in

this paper). Furthermore, if a student could write more ideas with fewer words in a summary, his/her terse value would be higher.

Results and discussion

The change of summaries

Table 4 shows the changes of the students' summaries after the game. Statistical tests showed that the students significantly improved their summaries in terms of the numbers of main ideas and words. In other words, after the game, the students became able to write more main ideas in longer summaries. However, the change of the terse value was not significant, suggesting that the conciseness of important information remain the same. The stable terse values also implied that these graduate students did not increase words to write unimportant information. Instead, after the game, they were able to notice more complete main ideas. This was probably due to the reason that keywords may highlight what they should pay attention to and reorganizes the gist (Chou, 2012).

Table 4. The change of summaries after the game

Variables	*M* (*SD*)		*t*	*SE*	*p*
	Prior summaries	Final summaries			
# of main ideas	1.406 (0.920)	1.836 (0.856)	2.396*	0.179	.023
# of words	85.625 (44.661)	128.906 (35.931)	5.133***	8.432	< .001
Terse value	0.015 (0.009)	0.014 (0.006)	0.711	0.002	.482

Note. $^{*}p < .05$; $^{***}p < .001$.

Table 5 further shows the completeness of each main idea before and after the game. Several *t* tests showed that the students significantly improved the completeness of their ideas about the research motivation and objective. However, their ideas about the hypotheses of independent variables and interaction were not significantly changed. It could be noted that, before the game, these graduate students did not particularly pay attention to research motivation and objective in the section of introduction. Instead, they focused on the explanation of each hypothesis and overlooked the whole picture. By evaluating the keywords in the game, the students could better organize their summaries.

Table 5. The change of the completeness of ideas after the game

Main ideas	*M* (*SD*)		*t*	*SE*	*p*
	Prior summaries	Final summaries			
Research motivation	0.938 (0.716)	1.281 (0.924)	2.156*	0.159	.039
Hypothesis A	1.813 (1.491)	2.156 (1.370)	1.187	0.290	.244
Hypothesis B	1.594 (1.521)	1.594 (1.521)	0.000	0.269	1.000
Hypothesis A×B	0.781 (0.792)	1.000 (1.107)	1.070	0.204	.293
Research objective	0.500 (0.950)	1.313 (1.355)	4.104***	0.198	< .001

Note. $^{*}p < .05$; $^{***}p < .001$.

These findings were basically consistent with the survey of Evan and Green (2007), in which Chinese students found their difficulties in determining the main ideas of academic articles. This study further revealed the reading problems of the first-year Chinese graduate students. Specifically, when the students read the section of the introduction in an academic paper, they usually overlooked the structure of the article. They focused too much on relatively unimportant details. As a result, they could not form a systematic summary to fairly describe the content. In this vein, they probably lacked the ability of structure analysis and idea integration. Before reading, students should look over the text first and note the text structure (Duke & Pearson, 2002). One of the solutions was the scaffold of concept maps (e.g., Sung et al., 2016). Such graphical representation might help students organize better summaries. On the other hand, although the students improved their descriptions of research motivation and objectives in this study, there was still much room for increasing the completeness of their summaries. For each main idea, most of the students put too much emphasis on describing the explanations and forgot the arguments. This could result from lacking the ability of argumentation. By scaffolding students to make arguments, they should be able to construct more logical summaries.

The relationships among the evaluation and usage of keywords

Table 6 shows the spearman correlation coefficients among the evaluation and usage of keywords. As shown in the table, the students' evaluation of keywords had significantly medium positive correlation with the evaluation of experts, suggesting that these graduate students in general were able to distinguish the importance of the keywords. Besides, the frequencies of the keywords in the students' prior summaries were also positively related to experts' evaluation. Although the frequencies of the keywords in prior summaries were associated with students' evaluation, the correlation was low. The reason may be that when the students composed summaries at the first time, they had to focus more on the original article rather than considering the importance of the keywords.

Table 6. The Spearman correlation coefficients among the evaluation and usage of keywords

Spearman correlation coefficients	(1)	(2)	(3)	(4)	(5)	(6)
(1) Expert evaluation	-					
(2) Student evaluation	0.509***	-				
(3) Frequencies in prior summaries	0.685***	0.396***	-			
(4) Bids offered	0.348***	0.596***	0.309***	-		
(5) Purchase	0.102	0.331***	0.077	0.618***	-	
(6) Frequencies in final summaries	0.185**	0.443***	0.181**	0.655***	0.810***	-

Note. $^{**}p < .01$; $^{***}p < .001$.

After the activity, the frequencies of the keywords in their final summaries had a lower correlation with experts' evaluation, and had a higher correlation with their own evaluation. The result suggested that they later became freer to use these keywords in their summaries. Interestingly, the behaviors in the game were also associated with these factors. Furthermore, there was a significant moderate positive correlation between students' evaluation and bids offered in the games, while there was a low correlation between their bids and experts' evaluation. The result suggested that they more likely offered high bids to their important keywords.

However, the behaviors of purchase were not necessarily related to these factors. Furthermore, the behaviors of purchase were mainly associated with students' evaluation and bids, but not associated with experts' evaluation or the usage in their prior summaries. This was probably because their purchase behaviors were constrained by the game rules. They could only make decisions to buy the items of their highest bids. Similarly, the frequencies of the keywords in the final summaries were highly related to what they had purchased.

On the other hand, students' explanation of their behaviors may reveal their opinions on this game. Some students indicated the reason why they liked this game. For example, S42 mentioned that, "there was high uncertainty in the auction, in which the players not only had to tell the importance of the words, but also had to guess the bids offered by the opponents." Similarly, S13 said, when bidding keywords, she "had to speculate the mental activities of other players, requiring players' multiple abilities." Another student S81 expressed his preference of strategy games, because "the mental competition may increase the playfulness." However, some students did not like the type of auction games. For example, S11 said that, she "had to buy her wanted keywords at a very high price," which was "less relevant to reading." These opinions suggested that the playfulness of the game was somehow related to its decision making process. Furthermore, when the students made the decisions to offer bids, they had to consider both whether the keywords had high values, and whether other players would offer high bids. The complexity of situation analysis required students' reasoning and speculation, likely resulted in playful thinking process. Unfortunately, the complexity could also be a two-edged sword. Some students who were not good at mental evaluation could feel frustrated and anxious. How to balance the complexity of games and playfulness will be an interesting issue to explore in the future.

Among all keywords, there were three keywords of which students' evaluation were significantly positively related to the bids. As shown in Table 7, they were "games," "load," and "interaction." The results suggested that when the students offered bids to the three keywords, they considered their importance to the article. Conversely, when offering bids to the other keywords, they considered other factors rather than the importance. The students reported that when they bidding keywords, they also considered "whether the keywords could be replaceable in addition to their importance (S11)," and "to spent as few as possible (S62)." If they could not purchase their wanted keywords, they may use "synonyms to replace the keywords (S42)" or "similar sentences to express the same meanings (S52)."

Previous researchers required students to generate keywords before summary writing (Chou, 2012), while this study allowed students to evaluate keywords and make decisions. Both methods may facilitate students to

improve their summaries. This may be due to the fact that composing short words to replace main ideas in articles was crucial to summary writing (see Duke, & Pearson, 2002). Furthermore, when summarizing articles, students were suggested to delete unnecessary information first, abbreviate main ideas as keywords, and finally invent topic sentence. Taking the advantage of games, this study provides students with opportunities to evaluate the importance of keywords instead of generating them. By doing so, they could mentally construct the relationship between keywords and summaries. By comparing the collective evaluation with their mental evaluation, the students had to make tough decisions. The results suggested that students' mental evaluation was correlated with their game behaviors and the usage of keywords in the summaries.

Table 7. The correlation between students' evaluation and bids offering to keywords

Keywords	Students' evaluation (1~4)	Students' bids (1~4)	Spearman correlation coefficients	*p*
Games	3.22 (0.66)	3.03 (0.95)	0.564**	0.001
Digital	3.04 (0.78)	2.16 (0.94)	0.089	0.629
Learning	3.16 (0.69)	2.44 (0.90)	0.052	0.777
Load	2.31 (0.81)	2.03 (0.98)	0.411*	0.019
Teaching	2.25 (0.92)	1.91 (0.98)	0.282	0.120
Interaction	2.55 (0.76)	2.31 (0.98)	0.392*	0.027
Internal	2.10 (0.94)	1.88 (1.05)	0.168	0.258
Hot issues	1.42 (0.70)	1.28 (0.62)	0.175	0.339

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

Conclusion

This study designed a summarization game by incorporating keyword selection and auction mechanism. Although summarization is essentially a personal learning activity, group-based games may expend students' horizon of reading. Furthermore, the auction mechanism provides students with opportunities to compare their evaluation of keywords with others, further stimulating students' way of thinking. In this vein, auction games, which are rarely adopted in education, provide students with opportunities to express and quantify their ideas. By doing so, they may examine their own understanding further. In a sense, the game limits students' selection of keywords, which allows them to make decision by evaluating the importance of keywords. However, even if students do not acquire important keywords, they still could replace them with synonyms or related words. More importantly, they may add more important ideas to compose summaries without addressing redundant or unnecessary information.

Most educational games focus on well-structured subject domains, probably because the knowledge in these subject domains is quantifiable. However, the knowledge in ill-structured subjects, like language, cannot be easily gamified, except basic facts. Instead, some researchers attempted to create a simulated, augmented, or authentic environment so that students can experience the use of languages in situated environments. The game rules of text games are relatively hard to design because games should fairly deal with human creations (e.g., text, speech, motions). This study provides a feasible example to start up such creation games. Furthermore, this game adopts a word segmentation technique to extract keywords. By weighting these keywords, one's evaluation becomes quantifiable and human creations can be fairly gamified.

Besides, this study not only demonstrates students' improvement of summary writing, but also explores the reading problems of graduate students. Specifically, they may lack the abilities of structure analysis, idea integration and argumentation. Unlike children, adult readers usually have specific intentions and prior knowledge. Even for first year graduate students, after a while, they likely have advanced academic needs, which may be supported by advanced technology. For example, after summarization, automatic essay scoring could be used to provide students with personal feedback for improving their writing. Alternative, their summaries could be compared with other players', so that they could do self-reflection. On the other hand, text mining could be applied to automatically retrieve expert knowledge from research papers given specific research issues. By doing so, learning activities for graduate education like iAbstract could be extendable to other issues, other subject domains or even other ages.

Limitations and future works

Because this was an exploratory study, there were several limitations in this paper. First, our study adopted the summarization approach of McNeil and Donant (1982) as the design framework. The iAbstract game is only an implemented example of the framework. There are still many feasible ways to gamifying the process of summarization, while many further issues will emerge in these new ways. Furthermore, the auction mechanism provides students with interactions among players to facilitate their evaluation of the information in articles and the composition of short words for creating topic sentences. Such interactions, which may invoke or hinder students' thinking, needs further investigation. Second, the game designed in this study initially incorporated some techniques of natural language processing to support the gameplay. Recently, the technologies of online text analysis become more mature. For example, Latent Dirichlet Allocation can be applied to dynamically analyze students' topics and emotions in games from students-generated texts. Third, this study lacked rigorous experimental design, which required at least a control group and longer time. For this reason, the results may not be overgeneralized to other contexts. The learning effects of this educational game are also not included in this study and need to be examined in the future.

Acknowledgements

This study was funded by the National Key Research and Development Program of China (No. 2018YFB1004504), Hubei Provincial Science and Technology Program of China (No. 2017AKA191), and Self-determined Research Funds of Central China Normal University from the Colleges Basic Research and Operation of Ministry of Education of China (No. CCNU16GD003).

References

Baker, W., & Boonkit, K. (2004). Learning strategies in reading and writing: EAP contexts. *RELC Journal*, *35*(3), 299-328.

Brown, J. S., Collins, A., & Duguid, P. (1989). Situated cognition and the culture of learning. *Educational researcher*, *18*(1), 32-42.

Carroll, D. W. (2008). Brief report: A Simple stimulus for student writing and learning in the introductory psychology course. *North American Journal of Psychology, 10*(1), 159-164.

Chang, W. C., & Ku, Y. M. (2015). The Effects of note-taking skills instruction on elementary students' reading. *The Journal of Educational Research*, *108*(4), 278-291.

Chen, I.-H., Yang, Y.-T., & Wen, D. X. (2016). The Study of flow experience in digital game-based learning. *Digital Education Research, 8*, 40-46. [Chinese Journal Paper]

Chou, M. H. (2012). Implementing keyword and question generation approaches in teaching EFL summary writing. *English Language Teaching*, *5*(12), 36-41.

Dempsey, K., Jackson, G. T., & McNamara, D. S. (2010). MiBoard: creating a virtual environment from a physical environment. In V. Aleven, J. Kay, & J. Mostow (Eds.), *Intelligent Tutoring Systems 2010, Part II* (pp. 294-296). Pittsburgh, PA: Springer-Verlag Berlin Heidelberg.

Duke, N. K., & Pearson, P. D. (2002). Effective practices for developing reading comprehension. In N. K. Duke, P. D. Pearson, A. E. Farstrup, & S. J. Samuels (Eds.), *What research has to say about reading instruction* (pp. 205-242). Newark, DE: International Reading Association.

Evans, S., & Green, C. (2007). Why EAP is necessary: A Survey of Hong Kong tertiary students. *Journal of English for Academic Purposes*, *6*(1), 3-17.

He, Y., Hui, S. C., & Quan, T. T. (2009). Automatic summary assessment for intelligent tutoring systems. *Computers & Education, 53*(3), 890-899.

Herrington, J., & Oliver, R. (2000). An Instructional design framework for authentic learning environments. *Educational Technology Research and Development*, *48*(3), 23-48.

Herrington, J., Reeves, T. C., & Oliver, R. (2014). Authentic learning environments. In J. M. Spector, M. D. Merrill, J. Elen, & M. J. Bishop (Eds.), *Handbook of Research on Educational Communications and Technology* (pp. 401-412). New York, NY: Springer.

Huang, C. S., Yang, S. J., Chiang, T. H., & Su, A. Y. (2016). Effects of situated mobile learning approach on learning motivation and performance of EFL students. *Journal of Educational Technology & Society*, *19*(1), 263-276.

Hwang, G. J., Chiu, L. Y., & Chen, C. H. (2015). A Contextual game-based learning approach to improving students' inquiry-based learning performance in social studies courses. *Computers & Education*, *81*, 13-25.

Kim, A. Y. A., & Kim, H. J. (2017). The Effectiveness of instructor feedback for learning-oriented language assessment: Using an integrated reading-to-write task for English for academic purposes. *Assessing Writing, 32*, 57-71.

Kintsch, E., Steinhart, D., Stahl, G., Matthews, C., Lamb, R., & the LSA Research Group. (2000). Developing summarization skills through the use of LSA-based feedback. *Interactive Learning Environments, 8*(2), 87-109.

Konuk, S., Ören, Z., Benzer, A., & Sefer, A. (2016). A Study on creating writing strategy and evaluation tool for book summary. *Educational Research and Reviews*, *11*(21), 2021-2033.

Landauer, T. K., Lochbaum, K. E., & Dooley, S. (2009). A New formative assessment technology for reading and writing. *Theory into Practice, 48*(1), 44-52.

Liu, J. Y., Chang, Y. J., Yang, F. Y., & Sun, Y. C. (2011). Is what I need what I want? Reconceptualising college students' needs in English courses for general and specific/academic purposes. *Journal of English for Academic Purposes*, *10*(4), 271-280.

Magliano, J. P., Todaro, S., Millis, K., Wiemer-Hastings, K., Kim, H. J., & McNamara, D. S. (2005). Changes in reading strategies as a function of reading training: A Comparison of live and computerized training. *Journal of Educational Computing Research*, *32*(2), 185-208.

Maina, F. W. (2004). Authentic learning: Perspectives from contemporary educators. *Journal of Authentic Learning, 1*(1), 1-8.

McNamara D. S., Levinstein I. B., & Boonthum C. (2004). iSTART: interactive strategy training for active reading and thinking. *Behavior Research Methods, 36*(2), 222-233.

McNamara, D. S. (2009). The Importance of teaching reading strategies. *Perspectives on Language and Literacy, 35*, 34-38.

McNeil, J., & Donant, L. (1982). Summarization strategy for improving reading comprehension. In J. A. Niles & L. A. Harris (Eds.), *New Inquiries in Reading Research and Instruction* (pp. 215-219). Rochester, NY: National Reading Conference.

Park, S. W., & Kim, C. M. (2015). The Effects of a virtual tutee system on academic reading engagement in a college classroom. *Educational Technology Research and Development, 64*(2), 195-218.

Shih, M. (1992). Beyond comprehension exercises in the ESL academic reading class. *TESOL Quarterly, 26*(2), 289-318.

Sung, Y. T., Liao, C. N., Chang, T. H., Chen, C. L., & Chang, K. E. (2016). The Effect of online summary assessment and feedback system on the summary writing on 6th graders: The LSA-based technique. *Computers & Education, 95*, 1-18.

Wade-Stein, D., & Kintsch, E. (2004). Summary street: Interactive computer support for writing. *Cognition and Instruction, 22*(3), 333-362.