A Vocal Approach in Measuring Critical Thinking Skill in Elementary School Students

Xuan-Yang HUANG, Sie Wai CHEW & Nian-Shing CHEN[*]
Department of Information Management, National Sun Yat-sen University, Kaohsiung, Taiwan
*nschen@mis.nsysu.edu.tw

Abstract: Critical thinking is an important skill in the 21st century. With the awareness of the importance of critical thinking skill, schools have invested much effort in including the development of critical thinking skill in their curriculum. Yet, there has not been much critical thinking assessments available to measure students' critical thinking skill as an independent skill, especially for elementary school students. Past research have venture into developing critical thinking skill assessments using children stories and folk stories by using different input methods (i.e., written and vocal assessment). However, there was a research gap in between these research where the effect of both input method on the critical thinking assessment was not explored. Hence, this research utilized a local folk story to design a critical thinking assessment, instead of using a conventional written assessment, a vocal assessment with a recording function was used to compare the difference of both assessments. The results indicates that there were no difference in the different input method on the effectiveness of the designed critical thinking assessment in assessing students' critical thinking skill. However, other interesting elements were found in this research, including a noticeable difference in the students' performance on certain questions, and the difference in the students' perception in terms of the difficulty of the questions and their intention to have their peers participate in the research.

Keywords: Critical thinking, assessment, vocal, written, elementary students

1. Introduction

Critical thinking is an important skill in 21st century, along with other skills such as collaboration, communication, and problem solving (Kay & Greenhill, 2011). Critical thinking skill is not something new as schools had been including them in their curriculum by applying higher level thinking skills of Bloom's taxonomy in their classes which includes analysis, synthesis and evaluation (Krathwohl, 2002; Lee et al., 2016). Although the importance of critical thinking to global education has been emphasized repeatedly in different avenues, not enough work has been done to measure these skills in a classroom setting, especially among elementary school students. Furthermore, with the importance of critical thinking skills increasing in the 21st century, an adequate critical thinking assessment become a major focus in recent research (Gelerstein, del Río, Nussbaum, Chiuminatto, & López, 2016; Lin et al., 2017).

According to UNESCO, measuring improvements in critical thinking skills is essential in providing a reference to an individual's progress in enhancing their critical thinking skill, and to better improve the quality of education provided to them (Ernst & Monroe, 2004). There are many critical thinking assessments available, such as The Ennis-Weir Critical Thinking Essay Test (Ennis & Weir, 1985), and the California Critical Thinking Skills Test (CCTST) by Insight Assessment (Tung & Chang, 2009). However, these assessments were designed mainly for high school students, college students, or adults. As Kuhn (1999) mentioned that young children have the cognitive ability to have opinions and ideas like adults, the cultivation of critical thinking skill should begin at a young age. Therefore, more research efforts could be invested in the development of critical thinking assessments suitable for elementary school students.

Past research have designed different critical thinking assessments for elementary school student, such as Gelerstein's et al. (2016) and Lin's et al. (2017) research that involved the usage of comic story to assess elementary students' critical thinking skills level. These research utilized a

different approach in assessing students' critical thinking skills. Instead of using conventional multiple choice questionnaires, these research used a mixture of close-ended and open-ended questions with questions designed in accordance to the definition of critical thinking skill by Facione (1990). However, these research utilized different input methods in supporting the students in answering the designed critical thinking assessment. In Gelerstein's et al. (2016) research, students were required to write down their answers on the provided assessment script. With the advancement of technology, in Lin's et al. (2017) research, a vocal critical thinking assessment was designed where students were required to speak up and had their answers recorded by the designed computer based. With the introduction of the vocal input and improved storytelling features in Lin's et al. (2017) research, as compared to Gelerstein's et al. (2016) research, it was noticed that there was a research gap in between these two research where the effect of the different input method was not studied.

Hence, the research objective of this study was to determine the effect of using different input methods (i.e., written and vocal) supported by technology on the same critical thinking assessment in assessing the critical thinking skill of elementary school students. The research questions of this study are "By using the same critical thinking assessment to assess the critical thinking skill of elementary school students, what are the effect of using different input methods in answering the assessment?" and "Does gender difference affect students' performance in the critical thinking assessment?"

2. Literature Review

Scholars have been defining critical thinking from different aspects. Ennis (2011) mentioned that "[c]ritical thinking is reasonable and reflective thinking focused on deciding what to believe or do" (p. 1). Harpern (2014) defined critical thinking as "the use of those cognitive skills or strategies that increase the probability of a desirable outcome" (p. 8). A critical thinker seeks for evidence and supporting argument when making decisions to avoid biasness and unfairness, resulting in a desirable outcome which may differ in the values of each individual (Halpern, 2014). Critical thinking is described as an attitude of the mind and its character to reply as it is a development of successful information-processing (Chen, Tolmie, & Wang, 2017). Facione (1990) provided a more structured definition of critical thinking where critical thinking is "purposeful, self-regulatory judgment which results in interpretation, analysis, evaluation, and inference, as well as explanation of the evidential, conceptual, methodological, criteriological, or contextual considerations upon which that judgment is based" (p. 6).

Based on Facione's (1990) Delphi report, critical thinking consist of "Interpretation," "Analysis," "Evaluation," "Inference," "Explanation" and "Self-regulation." "Interpretation" involved having one understanding the role and importance of an event or information represented. "Analysis" involved having one recognizing the inferential relationships among statements or concepts in describing the opinion or belief. "Evaluation" involved having one assessing the credibility of statements and assessing the reasoning provided. "Inference" involved having one using relevant information to deduce objective conclusions, form hypotheses, and understand the flowing consequences of the statement or event. "Explanation" involved having one describing the result of one's reasoning and to prove that reasoning behind the provided result. "Self-regulation" involved having one applying the abovementioned critical thinking skill on their own critical thinking thoughts and opinions (Facione, 1990). The definitions of each skill are discussed in the Delphi report by Facione (1990).

Past research have utilized Facione's (1990) definition of critical thinking in developing critical thinking skill assessments suitable for elementary school students (e.g., Gelerstein et al., 2016; Lin et al., 2017). These research utilized children stories and folk stories which were familiar to elementary school students in developing their critical thinking skill assessments. Past research have shown that critical thinking skill has strong connection between reading comprehension (Norris & Ennis, 1989; Tung & Chang, 2009). By introducing the mechanism of storytelling with critical thinking skill, it would improve the performance of the assessment and enhance students' experience in answering the assessment (Barker & Gower, 2010; Yang & Wu, 2012; Yeh, 2001). Furthermore, past research have proven that the cognitive activity including critical thinking would

had begun to develop at a young age, from the level of K-12, and maybe even earlier than kindergarten (Daniel, & Gagnon, 2011; Florea & Hurjui, 2015; Willingham, 2008; Kennedy, Fisher, & Ennis, 1991).

Past research have utilized difference answering methods on critical thinking assessments, such as in written form and in verbal form (Gelerstein et al., 2016; Lin et al., 2017). Past research that utilized a written critical thinking assessment in Taiwan had found that elementary school students may poses sufficient vocabulary to express their opinions, but when they were required to have them written down, these students would face difficulties in writing down their answers as they only begin to write the Chinese character at Grade 3 and Grade 4 (Chew et al., 2018). In additional to that, it was found that there were research that utilized only written assessments, with other research that utilized only vocal assessments in measuring students' critical thinking skill (Gelerstein et al., 2016; Lin et al., 2017). Yet, there have not been a research exploring the effects of using these two different input methods on the results of the critical thinking assessment. Hence, in order to explore this effect, by maintaining the storytelling features introduced in Lin's et al. (2017) research, the research objective of this study was to explore the effect of using different input methods (i.e., written and vocal) on the same critical thinking assessment in assessing the critical thinking skill of elementary school students.

3. Research Design

3.1 Participant

The target participants of this study were from an elementary school from Kaohsiung, Taiwan. Students for this study were fourth grade students, with a total of 100 students (43 male and 57 female). The students were randomly selected to be either complete the written assessment or complete the vocal assessment with 50 students assigned to complete the written assessment (22 male and 28 female) and another 50 assigned to complete the vocal assessment (21 male and 29 female).

3.2 Assessment Design

Past study have shown that Taiwanese elementary students might face difficulty in understanding Western stories (Lin et al., 2017). Hence, in this study, the critical thinking assessment was designed using a famous Chinese traditional story, "Chinese Zodiac Signs." Critical thinking questions were inserted in the stories, where the questions were designed based on the story plots. The story was selected because of the participants were from Taiwanese elementary school students who were familiar with the story.

Table 1
Distribution of Critical Thinking Skill Questions

Critical thinking skill	Number of questions
Interpretation	5
Analysis	5
Evaluation	4
Inference	4
Explanation	4
Total	22

After finalizing critical thinking assessment's context, next step was to design critical thinking questions in the assessment to measure students' critical thinking skills. The designed questions were based on Facione's (1990) definition of critical thinking skills which consist of six different dimensions, such as "Interpretation," "Analysis," "Evaluation," "Inference," "Explanation" and "Self-regulation." The skill of "Self-regulation" was not included in this study as the design of the study was not capable to support the assessment of this skill. The distribution of the designed 22 critical thinking questions by its designated skills for the story "Chinese Zodiac Signs"

were shown in Table 1. As an example, in the story the mouse and the cat was having trouble finding a way to cross the river. Students were required to suggest methods that could help the mouse and the cat cross the river where this question was designed to measure the "Inference" skill. The follow up question for this question required students to explain their reasons in supporting their suggestion made in assisting the mouse and the cat cross the river which was designed to measure the "Explanation" skill.

The designed critical thinking assessment consist of open ended questions, mix and match questions, and multiple-choice questions. A Q-sort analysis was conducted with two researchers to categorize the designed questions into each critical thinking skill in accordance to the definition of Facione (1990). Both researchers completed the Q-sort analysis independently. The Q-sort result from both researchers were compared with the study's initial categories of the designed questions. Thereafter, a meeting was conducted with each of the researchers separately in order to understand the reason of their categorizing results. Necessary changes and amendments were done in accordance to the feedback from both researchers to improve the questions in the critical thinking assessment. After finalizing the questions, the critical thinking assessment was reviewed by elementary school teacher to make sure students could understand the wordings used in the questions.

3.3 System Design

For this study, a computer based application was designed to include the storytelling process, and the question and answer process. The application used pictures portraying the scene of the story with written dialogues and narrative supported by recorded voice over for the narrative and each character in the story (see Figure 1(a)), which was similar with the application used in Lin's et al. (2017) research. The application was designed using Microsoft Visual Studio where the pictures and voice recording were aligned to enable to storytelling process. For the questions and answer process, the questions were displayed in wordings along with the recorded voice over. For closed ended questions, the application provided the relevant selections (see Question 9 in Figure 1(b)). For open ended questions, the application required students to record down their answers by clicking on the recording button, and to click on the stop recording once they had completed their answers (see Question 10 in Figure 1(b)). These questions were inserted in between different parts of the story. Once they had completed a question, the application continued on with the storytelling process until the next designed question came up.
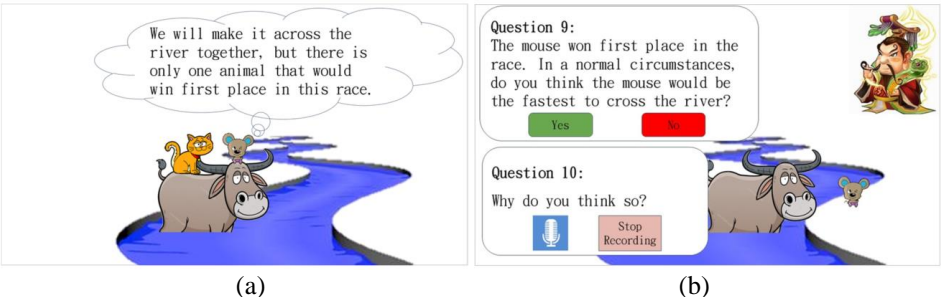

Figure 1. (a) Storytelling process, (b) Question interface of the vocal assessment group

Figure 2. (a) Question interface of the written assessment group, and (b) answer sheet

For the written assessment group, a slight modification of the application was made. Both groups were required to listen to the storytelling process using the application via the computer and headset provided. The difference between both groups was that the students of the vocal assessment group were required to record their answer, the written assessment group were required to write down their answer on the answer sheet provided, resulting in a slight modification on the question and answer interface for the written assessment group (as shown in Figure 2). The selection buttons and the recording buttons were removed, instead a written instruction was provided to guide students in writing down their answers on the answer sheet provided.

3.4 Research Process

The study required the usage of a Windows computer for operation of the application, along with headsets with microphones. Hence, the study was conducted during the students' computer class in school. Students were randomly assigned to complete the assessment in a written form or by vocal input. All students used the designed application for the storytelling process. The difference between two groups were during the question and answer process where students of the written assessment group answered by writing down their answers on the answer sheets provided, and students of the vocal assessment group answered by recording their answers through the application (see Figure 3).

For each session, a briefing was conducted before the students began using the application to inform students on items to expect during the study. The application started by having students selecting their basic information, included class, seat number, name and gender. After filling in these information, the application entered the story along with the critical thinking questions. The given time period for the study was 40 minutes. With the completion of the story, students were required to complete a questionnaire on their opinion on the application and their familiarity of the story, where there were a total of eight 5-Likert scale questions.

| (a) | (b) |

Figure 3. (a) Student completing the written assessment, (b) Student completing the vocal assessment

4. Data Analysis

The academic results from last semester of the students were collected to compare the academic performance of the students who completed the critical thinking assessment by writing their answers with the students who completed the assessment verbally. The collected academic subjects included Mandarin, Mathematics and Science. The average score of these results for both groups were compared using t-test. The results indicates that both groups were similar in term of their academic performance where $t = 0.265$, $p = .791$ (Written assessment: $n = 50$, $M = 92.54$, $SD = 4.09$, and Vocal assessment: $n = 50$, $M = 92.30$, $SD = 5.16$). Furthermore, a gender comparison was also conducted, where the t-test results showed that there were not significant difference between both genders as well where $t = -1.449$ and $p = .150$ (as shown in Table 2). This indicates that both groups were of similar level for further comparison.

After the data was collected, two researchers were assigned to evaluate the data received from both written and vocal critical thinking assessment. The two researchers were briefed on the requirements of each question along with the critical thinking skill in evaluation of the question. A guideline was prepared, listing the designated critical thinking skill and the requirement for each question. When the students' answer fulfill the requirement of the critical thinking skill designated

for each question, the score "1" was given for that question, else score "0" was given. In order to ensure that both researchers had common understanding in evaluating the data, a Kappa coefficient was calculated. The Kappa coefficient between both researchers were above 0.80 which was considered acceptable (Landis & Koch, 1977).

Table 2
t-test Results of the Students' Academic Results

Variable		n	Mean	SD	t	p
Group	Written	50	92.54	4.09	0.265	.791
	Vocal	50	92.30	5.16		
Gender	Female	57	93.00	3.55	-1.449	.150
	Male	43	91.65	5.72		

The critical thinking assessment had a total of 22 questions consisting questions covering the five critical thinking skills (i.e., Interpretation, Analysis, Evaluation, Inference and Explanation) with the distribution shown in Table 1. The *t*-test results comparing both groups were as shown in Table 3. There was no significant difference between both groups in the results of the critical thinking assessment. In terms of the critical thinking skills, there were not significant difference as well for all skills except for "Inference" skill where $t = -3.326$ and $p = .001$.

Table 3
t-test Results of the Critical Thinking Assessment by Group

Critical thinking skill	Group	n	Mean	SD	t	p
Total	Written	50	14.44	4.234	-1.417	.160
	Vocal	50	15.60	3.949		
Interpretation	Written	50	3.46	1.313	-0.176	.861
	Vocal	50	3.50	.931		
Analysis	Written	50	3.66	1.118	-0.961	.339
	Vocal	50	3.88	1.172		
Evaluation	Written	50	2.86	.990	-0.796	.428
	Vocal	50	3.02	1.020		
Inference	Written	50	2.46	1.164	-3.326	.001[**]
	Vocal	50	3.14	.857		
Explanation	Written	50	2.00	1.294	-.227	.821
	Vocal	50	2.06	1.346		

Note. [**]$p < .01$.

For further analysis, since there were no significant difference in the results of the critical thinking assessment by each critical thinking skill except for the "Inference" skill, a *t*-test analysis was conducted on the results of each "Inference" skill question. The four questions designed in the assessment for the "Inference" skill required students to draw reasonable conclusions on their decision to support or to oppose a statement or a description. For example, Question 9 in Figure 1(b), students were asked whether in a normal circumstances the mouse would be the fastest animal to cross the river (which was designed to measure the "Evaluation" skill), and they were required to share the reasons behind their decision which was designed to measure the "Inference" skill. Besides that, for the "Inference" skill, the designed questions required students to propose alternatives by using different methods or strategies in solving a problem. For example, in the story, the rabbit was just behind the tiger and students were required to suggest methods in helping the rabbit cross the river given that the tiger was just about to cross. It was shown in Table 4 that for this question, the vocal assessment group had performed significantly better than the written group. This was also observed for the "Inference" skill question designed that required students to arrive at a valid conclusion based on evidence regarding the reason why the dragon was late to arrive, the vocal assessment group had also performed significantly better than the written group (as shown in Table 4).

Table 4
t-test Results of the "Inference" Questions in the Critical Thinking Assessment

Inference question	Group	n	Mean	SD	t	p
Please propose other methods that could assist the cat and the mouse to cross the river.	Written	50	.80	.404	-0.793	.430
	Vocal	50	.86	.351		
Please share your reason why you agree or disagree that the mouse would be the fastest animal to cross the river.	Written	50	.68	.471	-1.888	.062
	Vocal	50	.84	.370		
Please share what you would do if you were the rabbit in this situation.	Written	50	.32	.471	-2.466	.015*
	Vocal	50	.56	.501		
Could you think of any reason why the mighty dragon would be late?	Written	50	.66	.479	-2.681	.009**
	Vocal	50	.88	.328		

Note. *p < .05; **p < .01.

Since there were no significant difference between the two groups (i.e., written and vocal input), a further analysis was conducted to determine whether there was difference in the results of the critical thinking assessment by gender. The results in Table 5 indicates that there were significant difference in between both genders regardless of the input method of the assessment. This was especially visible in questions on the "Inference" and "Explanation" skill where the questions designed for both skills were open-ended questions. For these questions, the female students performed better than the male students, even though both genders were similar in terms of their academic performance.

Table 5
t-test Results of the Critical Thinking Assessment by Gender

Critical thinking skill	Group	n	Mean	SD	t	p
Total	Male	43	13.79	3.839	-2.674	.009**
	Female	57	15.95	4.103		
Interpretation	Male	43	3.28	1.120	-1.552	.124
	Female	57	3.63	1.128		
Analysis	Male	43	3.65	1.232	-0.901	.370
	Female	57	3.86	1.076		
Evaluation	Male	43	2.81	.958	-1.092	.277
	Female	57	3.04	1.034		
Inference	Male	43	2.44	1.140	-3.015	.003**
	Female	57	3.07	.942		
Explanation	Male	43	1.60	1.218	-2.916	.004**
	Female	57	2.35	1.302		

Note. **p < .01.

For the results of the 5-Likert scale questionnaire, the results were shown in Table 6. There were a total of eight questions in the questionnaire. The results showed that for six questions in the questionnaire, there were no significant difference in the results between both groups. However, there were two questions which had significant difference between both groups (Question 5 and Question 7, as shown in Table 6).

Table 6
t-test Results of the Questionnaire

Question	Group	n	Mean	SD	t	p
1. Do you like this application?	Written	50	3.14	1.262	-0.784	.435
	Vocal	50	3.34	1.287		
2. Do you enjoy using this application?	Written	50	3.30	1.199	-0.564	.574
	Vocal	50	3.44	1.280		
3. Do you find this application interesting?	Written	50	3.42	1.263	-0.318	.751
	Vocal	50	3.50	1.249		

4. Was this application easy to use?	Written	50	3.66	1.189	1.571	.119
	Vocal	50	3.26	1.352		
5. Would you invite your friends to use this application?	Written	50	2.64	.964	-2.321	.022[*]
	Vocal	50	3.12	1.100		
6. Would you like to use applications similar as this one?	Written	50	2.90	1.249	-1.172	.244
	Vocal	50	3.20	1.309		
7. Do you find the questions in the application difficult?	Written	50	2.72	.904	2.116	.037[*]
	Vocal	50	3.20	1.325		
8. How familiar are you with the story "Chinese Zodiac Signs"?	Written	50	4.18	1.137	0.173	.863
	Vocal	50	4.14	1.178		

Note. [*]$p < .05$.

5. Discussion

The study's research objective was to compare the effect of elementary school students using different input methods in answering the critical thinking assessment. The results in Table 3 indicates that there were no significant difference in the total results of the critical thinking assessment between the two groups. This indicates that both written and vocal critical thinking assessment could be used to assess the critical thinking skills of elementary school students.

With further analysis on the results by each critical thinking skill, it was found that except for the "Inference" skill, there were no significant difference between the two groups on the questions for other critical thinking skills (see Table 3). This may be due to the fact that all the "Inference" skill questions were open-ended questions that required student to justify and elaborate reason to support their decision, and to propose alternative methods to resolve a problem, as compared to the questions of other critical thinking skills. For other critical thinking skill, most of the questions designed were closed ended questions (such as, "Interpretation" and "Evaluation"). For questions on the "Explanation" skill, students were required to reasoning in justifying their view point, explaining their comprehension of a concept, and describing the strategy used in making their decision. For questions on the "Inference" skill, students were required to thinking of alternative methods, determine what information would be useful to support their argument, and deciding in between the advantages and disadvantages of a decision. These questions would require students to bring in new ideas, justify its importance in solving a problem, and to argue its relevance in solving a problem. Hence, with the additional cognitive activity required by the "Inference" questions, this may explain the reason resulting in the difference between the two groups of questions on "Inference" skill. Students of the vocal assessment group could speak up freely in answering the questions, as speech is a free form of expressing their thoughts and reasoning. On the other hand, for the written assessment group, by having to write down their answers, students would require a better structure in their thoughts and having their reasoning well organized to enable them to write down their answers.

As there were no significant difference between both of the groups, hence a further test was conducted to determine whether there were gender differences in the results of the critical thinking assessment. Differing from the results of Lin et al. (2017), this study found that there were significant difference between both genders even though there were no significant difference in their academic performance (see Table 5). The effect was evident on the designed questions for "Inference" and "Explanation" skills where questions were open-ended questions which required students to bring in new ideas and share their reasons to justify their answers. This effect could be due to the motivation and eagerness of the students when they were answering the critical thinking assessment. Through observation during the research, it was found that female students were more diligent in completing the critical thinking assessment as compared to the male students. This was also reflected in the data received as female students tended to be more meticulous in their answers as compared to the male students. This results was further supported by Lim, Han, Uhlhaas, and Kaiser's (2013) research finding that female students' brain tend to mature faster than male students of their age. The effect may be more evident in this research as male students as compared to Lin et al. (2017).

The results of the questionnaire, it was found that students of both groups enjoyed the designed application, and found that the application was easy to use and interesting. There were significant difference in between both groups in terms of the difficulty of the questions and whether would they invite their peers to try out this application. Although there were no significant difference in term of the results of the critical thinking assessment, the results of the questionnaire indicates that students of the written assessment group would not recommend this application to their peers as compared to the vocal assessment group. This may be due to the extra effort required in writing down their answers on the answer sheet, which was perceived as similar with additional homework. The vocal assessment group scored significantly higher than the written assessment group as the vocal assessment group was not required to write. Furthermore, this was supported by the question on whether the question in the application difficult, with the same questions applied on both groups, students of the written assessment group found the questions more difficult as compare to the vocal assessment group.

6. Conclusion

In conclusion, this study had found that both written and verbal critical thinking assessment could be used to assess the critical thinking skills of elementary school students, with the difference that more effort would be required for students to write down their answers for open-ended questions as compared to by saying their answers out loud. There were difference in the performance of the critical thinking assessment by gender where female students had significantly higher score as compared to the male students, given that both their academic performance were similar. This may be due to several factors including female students being more diligent in completing the assessment as compared to the male students. Another reason could be that female students' brain development tend to mature earlier than male students which would result in a better performance in the critical thinking assessment.

For future research, it is proposed that in additional of assessing the questions of the critical thinking assessment by a "0" and "1" manner (similar with Gelerstein et al., 2016 and Lin et al., 2017), the critical thinking indicator proposed by Newman et al. (1997) could also be adopted in assessing the critical thinking skills of students through content analysis. As Newman et al. (1997) suggested in their research, there are indicators of critical thinking skill such as Novelty, Importance, Relevance, etc. which are important elements of critical thinking. By using this approach to evaluate students' answer in these critical thinking assessment, it would provide another perspective on the critical thinking skills of students by different indicators, such as clarification, assessment (credibility), novelty, justification, linking idea, and importance (Lee et al., 2016).

Furthermore, with the usage of the recording function instead of a written assessment, students tend to be more willing to express their thoughts, although they may not use a complete sentence. For written assessments, students are accustomed to answer in a complete sentence as they were taught in class, which may hinder students in expressing their thoughts fully. With the recording mechanism, students could also practice their speaking skills and expressing their thoughts and ideas in a verbal manner. As "Self-regulation" was not monitored in this study, but with the recording function and the playback function incorporated in the designed application, the application could support "Self-regulation." With the proper guidance provided, this exercise would enable students to evaluate their own answers critically, allowing them to self-examine and self-correct. Furthermore, collaborative learning settings could be considered, where students could work as a group in evaluating their answers, and listening to different prospective from peers on the same question.

References

Barker, R. T., & Gower, K. (2010). Strategic application of storytelling in organizations: Toward effective communication in a diverse world. *The Journal of Business Communication (1973), 47*(3), 295-312.

Chen, P., Tolmie, A. K., & Wang, H. (2017). Growing the critical thinking of schoolchildren in Taiwan using the Analects of Confucius. *International Journal of Educational Research, 84*, 43-54.

Chew, S. W., Gelerstein, D., Lin, I.-H., Chen, C.-H., Chang, C.-M., Tsai, C.-C., & Chen, N.-S. (2018). *Comparing the Critical Thinking Skills among Elementary School Students in Taiwan and Chile.* Manuscript submitted for publication.

Daniel, M. F., & Gagnon, M. (2011). Developmental process of dialogical critical thinking in groups of pupils aged 4 to 12 years. Creative Education, *2*(05), 418-428.

Ennis, R. H. (2011). The nature of critical thinking: An outline of critical thinking dispositions and abilities. In *Sixth International Conference on Thinking, Cambridge, MA*.

Ennis, R. H., & Weir, E. E. (1985). *The Ennis-Weir critical thinking essay test: An instrument for teaching and testing*. Midwest Publications

Ernst, J., & Monroe, M. (2004). The effects of environment-based education on students' critical thinking skills and disposition toward critical thinking. *Environmental Education Research, 10*(4), 507-522.

Facione, P. (1990). *Critical thinking: A statement of expert consensus for purposes of educational assessment and instruction* (The Delphi Report).

Florea, N. M., & Hurjui, E. (2015). Critical thinking in elementary school children. *Procedia-Social and behavioral sciences, 180*, 565-572.

Gelerstein, D., del Río, R., Nussbaum, M., Chiuminatto, P., & López, X. (2016). Designing and implementing a test for measuring critical thinking in primary school. *Thinking Skills and Creativity, 20*, 40-49.

Halpern, D. (2014). *Thought and knowledge: An introduction to critical thinking* (5th ed.). Mahwah, NJ: Lawrence Erlbaum Associates.

Kay, K., & Greenhill, V. (2011). Twenty-first century students need 21st century skills. In *Bringing schools into the 21st century* (pp. 41-65). Springer, Dordrecht.

Kennedy, M., Fisher, M. B., & Ennis, R. H. (1991). Critical thinking: Literature review and needed research. *Educational values and cognitive instruction: Implications for reform, 2*, 11-40.

Krathwohl, D. R. (2002). A revision of Bloom's taxonomy: An overview. *Theory into Practice, 41*(4), 212e218.

Landis, J. R., & Koch, G. G. (1977). An application of hierarchical kappa-type statistics in the assessment of majority agreement among multiple observers. *Biometrics*, 363-374.

Lee, H., Parsons, D., Kwon, G., Kim, J., Petrova, K., Jeong, E., & Ryu, H. (2016). Cooperation begins: encouraging critical thinking skills through cooperative reciprocity using a mobile learning game. *Computers & Education, 97*, 97-115.

Lim, S., Han, C. E., Uhlhaas, P. J., & Kaiser, M. (2013). Preferential detachment during human brain development: age-and sex-specific structural connectivity in diffusion tensor imaging (DTI) data. *Cerebral Cortex, 25*(6), 1477-1489.

Lin, I.-H., Chew, S. W., & Chen, N.-S. (2017). A Vocal assessment approach to measure elementary school students' critical thinking skills. In *Proceedings of 2017 IEEE 17th International Conference on Advanced Learning Technologies (ICALT2017)* (pp. 419-421). Conference Publishing Services. doi:10.1109/ICALT.2017.30

Newman, D. R., Johnson, C., Webb, B., & Cochrane, C. (1997). Evaluating the quality of learning in computer supported co-operative learning. *Journal of the American Society for Information Science, 48*(6), 484-495.

Norris, S. P., & Ennis, R. H. (1989). What is critical thinking? *The practitioner's guide to teaching thinking series: Evaluating critical thinking* (pp. 1-26).

Tung, C. A., & Chang, S. Y. (2009). Developing critical thinking through literature reading. *Feng Chia Journal of Humanities and Social Sciences, 19*, 287-317

Willingham, D. T. (2008). Critical thinking: Why is it so hard to teach? *Arts Education Policy Review, 109*(4), 21-32.

Yang, Y. T. C., & Wu, W. C. I. (2012). Digital storytelling for enhancing student academic achievement, critical thinking, and learning motivation: A year-long experimental study. *Computers & education, 59*(2), 339-352.

Yeh, S. S. (2001). Tests worth teaching to: Constructing state-mandated tests that emphasize critical thinking. *Educational Researcher, 30*(9), 12-17.