教育研究與發展期刊　第十八卷第一期　2022 年 3 月　頁 107-127

DOI　10.6925/SCJ.202203_18(1).0004

Improving Knowledge Retention in a Systems Programming Course Through Confirmatory Multiple-Choice Questions

Chien-Hung Lai
Assistant Professor, Department of Information and Computer Engineering, Chung Yuan Christian University

Bin-Shyan Jong
Professor, Department of Information and Computer Engineering, Chung Yuan Christian University

Yen-Teh Hsia
Professor, Department of Information and Computer Engineering, Chung Yuan Christian University

Tsong-Wuu Lin
Professor, Department of Computer Science and Information Management, Soochow University

Abstract

Multiple-choice questions (MCQs) are often used in assessments. Although MCQs can be designed to test higher-order cognitive skills such as analysis, synthesis, and evaluation, they are often used to measure recall. This may encourage students to only learn material superficially. To resolve this problem, several variations of MCQs have been proposed to improve learning. We proposed a modified type of MCQ called a “confirmatory MCQ” (CMCQ). Optional online CMCQ tests were more effective than MCQs and confidence-based marking tests in helping students study material from a systems programming course because they prompted students to thoroughly evaluate each answer.

Keywords: multiple choice, knowledge retention, assessment, evaluation methodology, learning strategy

Chien-Hung Lai’s E-mail: soulwind@cycu.org.tw (Corresponding Author)

應用信心指標選擇題以保留學習知識：以系統程式課程為例

賴建宏 中原大學資訊工程學系助理教授
鍾斌賢 中原大學資訊工程學系教授
夏延德 中原大學資訊工程學系教授
林聰武 東吳大學資訊管理學系教授

摘　要

選擇題通常用於評估之目的。儘管根據先前研究選擇題被設計為可測試及評估更高階思維之認知處理，例如分析、綜合和評估，但多數時候選擇題仍然被使用於回測對於知識的短暫記憶保留，導致學習者為了達到良好的學習成就而採取死記硬背的方式，無法確實將知識長期記憶。有鑑於選擇題所擁有的潛在缺點，因此不少研究提出了改善選擇題的方式，而這些研究皆是為了讓學習者進行選擇題測驗，於進行形成性評估時能有更好的學習效果。本研究中提出了另一種的選擇題模式，稱為信心指標選擇題（confirmatory multiple-choice question, CMCQ），即測驗過程需給予每一選項信心指標，以表示對於該選項的確認程度，因此學生面對每一選項皆要全盤思考，將腦中學習過的課程知識整合，方能作出正確答案。同時為了檢測本研究設計能有效達到知識保留，讓學生在作答過程可以回顧先前學習過的內容，因此與一般的選擇題測驗及僅給予作答結果信心指標的選擇題測驗（confidence-based marking multiple-choice question, CBM-Based MCQ）進行比較，實驗結果顯示使用信心指標的選擇題測驗複習課程內容可以顯著幫助學習者更有效記住課程內容，這也是因為相較於回答一般選擇題或 CBM-Based MCQ 相比，對於每個選項都要確認信心指標的 CMCQ 更能刺激學生進行更徹底的思考，對學習過的內容反思，以達到知識保留。

關鍵詞：選擇題、知識保留、測驗評估、評估方法、學習策略

【通訊作者】賴建宏　電子郵件：soulwind@cycu.org.tw

I. Introduction

A. Research Background

Multiple-choice questions (MCQs) are often used in assessments (Tsai et al., 2015) and can be designed to test higher-order cognitive processes such as analysis, synthesis, and evaluation (Snyder & Snyder, 2008). However, MCQ tests may have disadvantages such as encouraging rote learning (Huang et al., 2018), measuring recall alone (Pamphlett & Farnill, 1995; Paxton, 2000; Tarrant et al., 2006), and encouraging surface learning (Scouller, 1998), which result in students only remembering answers without deeply understanding the subject matter; the design of teachers' questions may also contribute to these outcomes. Students answer memory-based questions by using only their memory. Thus, students prepare for exams mainly through pure memorization and ignore application, analysis, and integration (Candel et al., 2020). This may prevent students from learning to integrate newly acquired knowledge into their prior knowledge and lead them to only partially acquire new knowledge; consequently, they quickly forget the material (Tian et al., 2019). In meaningful learning, students must fully understand the concepts they learn. This results in long-term knowledge retention (Holley & Dansereau, 2014).

Several modifications to MCQs have been proposed to improve learning, namely using clickers, also known as "classroom response systems" (Fies & Marshall, 2006; Mayer et al., 2009), using hot designations, which prompt learners to think thoroughly when answering MCQs (Hsia et al., 2019), and employing multiple-choice concept maps (Novak, 1990; Sas, 2010), which test a learner's connected understanding (Ruiz-Primo et al., 2001a; Ruiz-Primo et al., 2001b; Schau et al., 2001). Studies have also explored confidence-based marking (CBM) of MCQs (Lai et al., 2014), which involves students indicating their confidence in their answers; rewards or penalties are assessed accordingly. CBM of MCQs requires students to select from three levels of confidence. In contrast to CBM of MCQs, reflective MCQs (RMCQs) require students to state their reasons for their answers (Ang & Boo, 2006). This idea is pedagogically sound and worth consideration because the main feature of MCQs is that test takers must select answers from available options, which can encourage students to guess; this also enables MCQs to be automatically scored.

MCQs should be modified to encourage students to reflect on their answers. Lai et al. (2014) proposed that MCQs can facilitate this process because students must select their level of confidence, which is based on reasoning. However, students may not carefully evaluate their answers. Students should evaluate every proposition implicit in MCQs before answering. The cognitive processes involved may help students clarify their thoughts by encouraging them to recall knowledge to support or refute the propositions (i.e., reasoning). However, students often answer MCQs by comparing the options and selecting the most promising one, especially if they do not know the answer immediately.

The following is an example MCQ:

- Who was sworn in as the US president when President Franklin D. Roosevelt passed away before the end of his fourth term?
  (1) Dwight D. Eisenhower
  (2) Gerald Ford
  (3) Lyndon B. Johnson
  (4) Harry S. Truman

This MCQ comprises four propositions: (a) Dwight D. Eisenhower was sworn in, (b) Gerald Ford was sworn in, (c) Lyndon B. Johnson was sworn in, and (d) Harry S. Truman was sworn in. The process of comparing the options to determine the most promising one differs from that of evaluating the truth of each proposition. Each proposition has a unique context with respect to which truth can be evaluated. When each proposition is considered in isolation, students can gather evidence to support or reject each on the basis of their contextual knowledge.

This study proposed that a novel form of MCQ, a confirmatory MCQ (CMCQ), can be used to review lessons. To answer CMCQs, students must consider each proposition individually and make a subjective, probabilistic estimation of whether it holds. In the example MCQ, President Roosevelt’s relation to President Eisenhower, for example, can be considered in isolation from the other propositions. If students attempt to recall their knowledge of all five US presidents (and their relationships) at once, they may overlook aspects they would otherwise notice were they to only recall their knowledge of two presidents (e.g., President Roosevelt and President Eisenhower). This is a tradeoff between time (to answer the question) and thoroughness (consideration). CMCQs apply to this tradeoff. When students answer CMCQs, they must make a subjective, probabilistic estimation of the truth of each proposition (Figure 1). Thus, CMCQs may help students to review their lessons and improve their learning process.

Figure 1
CMCQ

Who was sworn in as the next U.S. president when President Franklin D.Roosevelt (FDR) passed away before the end of his fourth term?

	Is definitely correct	**Is probably correct**	**No idea**	**Is probably incorrect**	**Is definitely incorrect**
(a) Dwight D. Eisenhower	☐	☐	☐	☐	☐
(b) Gerald Ford	☐	☐	☐	☐	☐
(c) Lyndon B. Johnson	☐	☐	☐	☐	☐
(d) Harry S. Truman	☐	☐	☐	☐	☐

B. Research Questions

To determine how CMCQs affect learning, this study conducted an educational experiment. Formative assessments in the form of MCQ tests (Tsai et al., 2015), CBM-based MCQ tests (Lai et al., 2014), and CMCQ tests were administered to help students understand the material covered in a systems programming course in the computer science department of a university in northern Taiwan. The students were randomly divided into traditional MCQ, CBM, and CMCQ groups. The traditional MCQ group was tested using MCQs (Figure 2), the CBM group was tested using CBM-based MCQs (Lai et al., 2014; Figure. 3), and the CMCQ group was tested using CMCQs (Figure 4).

This study (a) determined which group exhibited superior performance, (b) explored the differences in learning effects among the groups, (c) identified the reasons for these differences, and (d) examined the students' perspectives on CMCQs.

Figure 2

Practice Interface for Traditional MCQ Group

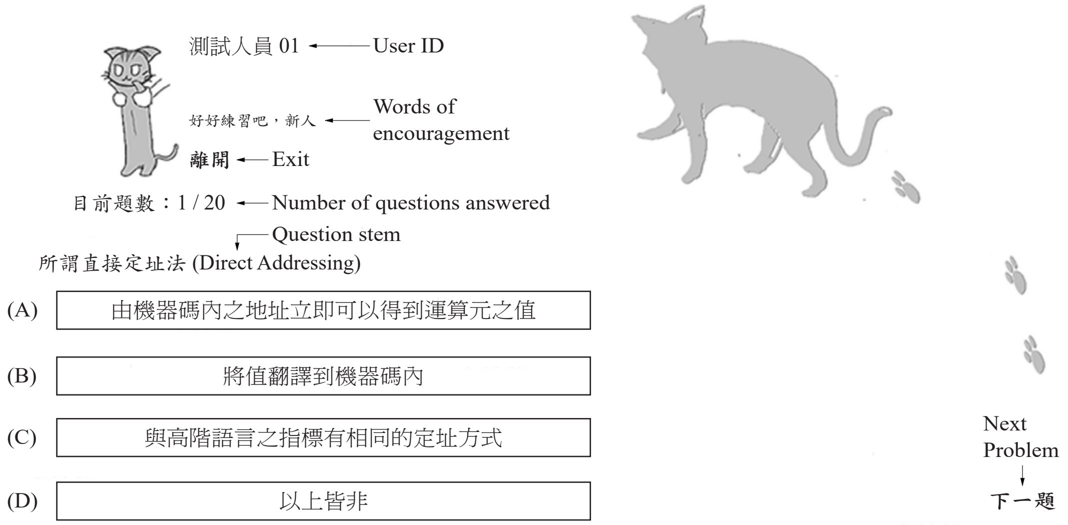


Figure 3

Practice Interface for CBM Group

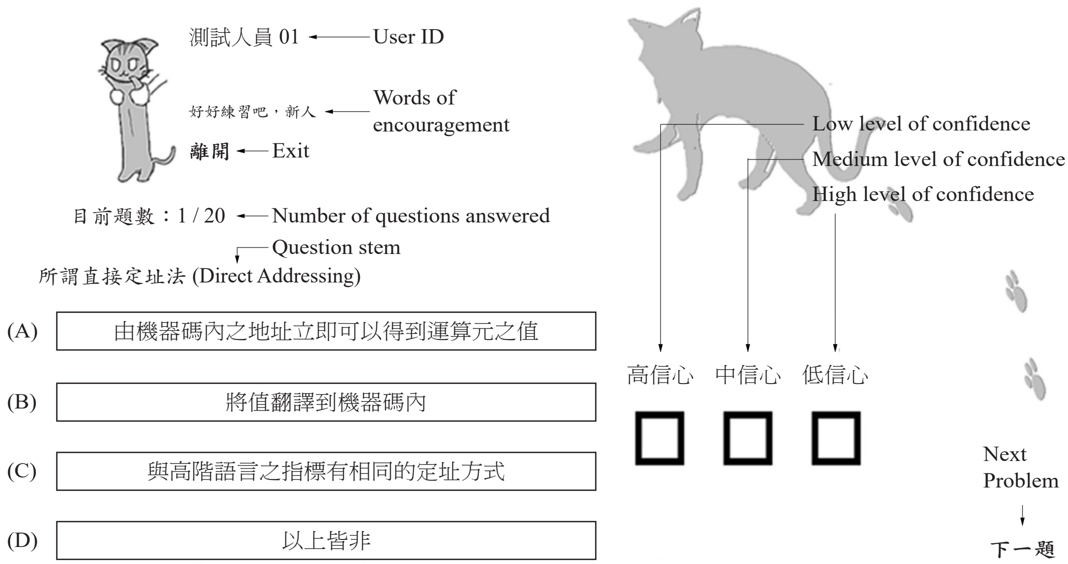
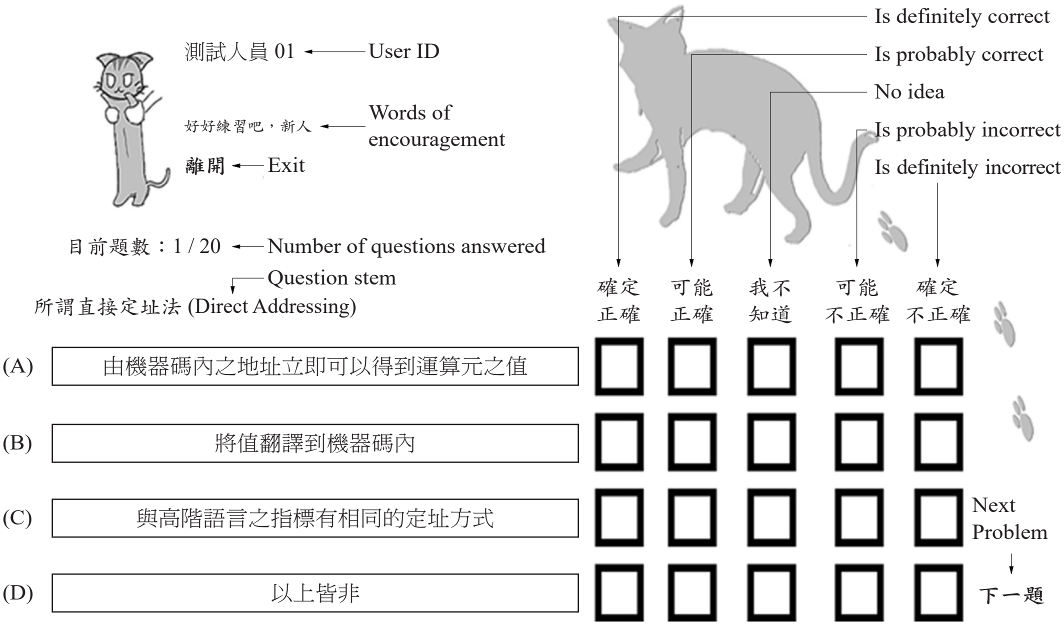


Figure 4
Practice Interface for CMCQ



II. Research Methods

A. Game-Based Review System

The authors developed a game-based review system to help the students review their lessons and increase their learning motivation, which increases learning efficiency. The students reviewed at their own pace and collected badges with cartoons on them. The students were able to access the practice interfaces and review the material at any time. The interface also maintained records of the students' answers and a leader scoreboard listing all students in order of the points they earned. With the sole exception of the practice interfaces, all treatments were identical among the groups. The only independent variable was the type of question (MCQs, CBM-based MCQs, and CMCQs) the students encountered in the interfaces.

The practice interface for the CBM group required students to select confidence levels (high, medium, or low; Figure 3). In the CMCQ group (Figure 4), students made probabilistic estimations (*definitely correct*, *probably correct*, *no idea*, *probably incorrect*, or *definitely incorrect*) for each option. The CMCQs were designed so that *definitely correct* could only be selected for one option and *definitely incorrect* could not be selected for all options.

B. Procedure

Before the experiment, the students took a pretest and were randomly divided into the three groups on the basis of their scores. On the first day of the experiment, the students were taught how to use the game-based system to review the material. The students then used the system in their free time to study for 4 weeks. On the last day of the experiment, the students took a posttest and completed a questionnaire regarding the system. The traditional MCQ and CBM groups also completed a subjective estimation (SE) questionnaire to assess the likelihood that they identified incorrect distractors under different circumstances. The SE consisted of eight items, each describing a distinct situation (e.g., "Suppose that after consideration, you felt that B was the answer and thought A, C, and D were wrong but were not sure. Would you review A, C, and D to ensure they were incorrect before clicking 'Next Problem' and submitting your answer?"), with six levels of probability (*definitely*, *most likely*, *probably*, *probably not*, *most likely not*, and *definitely not*). For each item on the SE questionnaire, the students offered a subjective assessment of the frequency of this scenario arising: *always*, *often*, *sometimes*, *not often*, *rarely*, and *never*.

One month after the experiment, all students completed a delayed posttest and two one-item questionnaires: the STRATEGY and UNDERSTANDING questionnaires. The STRATEGY questionnaire featured the question "When I answered an MCQ (or CBM-based MCQ or CMCQ) during practice, it forced me to think clearly and to confirm that the options I did not select were incorrect"; answers were given using a 7-point Likert-type scale. The UNDERSTANDING questionnaire featured the question "After I answered an MCQ (or CBM-based MCQ or CMCQ) during practice, I understood more about relevant issues," which was evaluated on a 7-point Likert-type scale.

III. Results

A. Overall Learning Effects

Table 1 presents the students' scores on the three exams. Table 2 presents the analysis of variance (ANOVA) results for the pretest and the analysis of covariance (ANCOVA) results for the immediate posttest (with the results from the pretest as the covariate) and delayed posttest (with the results of the pretest as the covariate). For the immediate posttest, no significant differences were observed between the two groups of each pair. For the delayed posttest, the CMCQ group outperformed both the CBM (F = 6.17, p < .05, d = 0.49) and traditional MCQ groups (F = 4.86, p < 0.05, d = 0.51).

Table 1
Exam Scores

Measurement	Group	n	M	SD
Pretest exam	Confirmatory	30	46.17	16.64
	Confidence	31	48.39	15.02
	Traditional	31	46.94	10.22
Immediate post-test exam	Confirmatory	30	65.03	17.69
	Confidence	31	65.65	18.11
	Traditional	31	58.94	19.62
Delayed post-test exam	Confirmatory	30	65.77	15.96
	Confidence	31	58.16	14.86
	Traditional	31	57.06	18.30

Table 2

ANOVA and ANCOVA Results for Exam Scores

ANOVA					
	SS	DF	MS	F	P
Between Groups	77.825	2	38.912	.193	.825
In the Group	17937.932	89	201.544		

ANOVA					
Measurement	Group		p	F	d
Pretest exam	Confirmatory	Confidence	.543	0.30	-0.14
	Confirmatory	Traditional	.833	0.83	-0.06
	Confidence	Traditional	.688	0.20	0.11

ANCOVA					
Measurement	Group		p	F	d
Immediate post-test exam	Confirmatory	Confidence	.882	0.02	-0.03
	Confirmatory	Traditional	.131	2.35	0.33
	Confidence	Traditional	.187	1.78	0.36
Delayed post-test exam	Confirmatory	Confidence	.016	6.17*	0.49
	Confirmatory	Traditional	.032	4.86*	0.51
	Confidence	Traditional	.994	0.01	0.07

Note. *p < .05

B. ANOVA Results for STRATEGY Ratings, UNDERSTANDING Ratings, and Answer Time

Table 3 presents the STRATEGY ratings, UNDERSTANDING ratings, and answer time (ANS-TIME) for the groups; Table 4 presents the ANOVA results for these measurements. Significant differences ($p < 0.001$) in STRATEGY and UNDERSTANDING ratings were observed between the CMCQ and CBM groups, and the differences were large ($d \geq 1.14$). Significant differences ($p < 0.001$) in STRATEGY and UNDERSTANDING ratings were also observed between the CMCQ and traditional MCQ groups, and the differences were even larger ($d \geq 1.46$). The UNDERSTANDING ratings of the CBM group were higher than those of the traditional MCQ group ($F = 6.52$, $p < 0.05$, $d = 0.64$). The ANS-TIME of the CMCQ group was longer than that of the traditional MCQ group ($F = 8.27$, $p < 0.01$, $d = 0.74$).

Table 3

STRATEGY Ratings, UNDERSTANDING Ratings, and ANS-TIME

Measurement	Group	n	M	SD
STRATEGY	Confirmatory	30	5.73	0.83
	Confidence	31	4.51	0.89
	Traditional	31	4.26	1.15
UNDERSTANDING	Confirmatory	30	5.63	0.96
	Confidence	31	4.55	0.93
	Traditional	31	3.81	1.33
ANS-TIME	Confirmatory	30	16.72	3.60
	Confidence	31	15.10	4.62
	Traditional	31	13.80	4.30

Table 4

ANOVA Results for STRATEGY Ratings, UNDERSTANDING Ratings, and ANS-TIME

Measurement	Group		F	d
STRATEGY	Confirmatory	Confidence	30.56***	1.42
	Confirmatory	Traditional	32.74***	1.46
	Confidence	Traditional	0.97	0.24

(Continued)

Table 4

ANOVA Results for STRATEGY Ratings, UNDERSTANDING Ratings, and ANS-TIME (continued)

Measurement	Group		F	d
UNDERSTANDING	Confirmatory	Confidence	20.11***	1.14
	Confirmatory	Traditional	37.62***	1.56
	Confidence	Traditional	6.52*	0.64
ANS-TIME	Confirmatory	Confidence	2.33	0.39
	Confirmatory	Traditional	8.27**	0.74
	Confidence	Traditional	1.32	0.29

Note. *$p < .05$; **$p < .01$; ***$p < .001$

C. Correlations

Table 5 presents Pearson correlations among the immediate posttest scores, delayed posttest scores, STRATEGY ratings, UNDERSTANDING ratings, and ANS-TIME. The STRATEGY ratings were highly correlated (r = 0.75) with ANS-TIME and moderately correlated (r = 0.22) with delayed posttest scores. A moderate correlation (r = 0.69) was also observed between STRATEGY and UNDERSTANDING ratings. In addition, moderate correlations were observed between ANS-TIME and UNDERSTANDING ratings (r = 0.51) and between immediate posttest and delayed posttest scores (r = 0.68).

The probability that the CBM and traditional MCQ groups confirmed that the distractors were incorrect (i.e., the SE value) was calculated (with the required normalization) by assigning numerical values to the SE answers [e.g., *definitely* (*always*) was 1, and *most likely* (*often*) was 0.8]. For both the traditional MCQ and CBM groups, the SE values were strongly correlated (r = .72–.73) with STRATEGY ratings (Table 6), which indicates the validity of the STRATEGY ratings.

Table 5

Correlations (Pearson's r*) between Exam Scores, Questionnaire Results, and ANS-TIME (*n *= 91)*

	IMMEDIATE	DELAYED	STRATEGY	UNDERSTANDING	ANS-TIME
IMMDIATE	-				
DELAYED	0.69***	-			
STRATEGY	0.08	0.22*	-		
UNDERSTANDING	0.12	0.14	0.69***	-	
ANS-TIME	-0.01	0.02	0.75***	0.51***	-

Note. IMMEDIATE = immediate posttest; DELAYED = delayed posttest exam; *$p < .05$; **$p < .01$; ***$p < .001$.

Table 6

Probability That Students Confirmed Distractors Were Incorrect and Correlation (Pearson's r*) with STRATEGY Ratings*

Measurement	Group	n	M	SD	Correlation with STRATEGY ratings
SE	Confidence	31	63.42	15.84	0.72**
	Traditional	31	59.01	11.33	0.73**

Note. **$p < 0.01$, M: average score (unit: points)

D. Student Feedback

Table 7 presents comments representative of the students' feedback. Two students from the traditional MCQ group indicated that the game-based system helped them study. Two students from the CBM group mentioned having a stronger impression of the answers they selected in answering CBM-based MCQs. Three students from the CMCQ group noted that answering CMCQs required thorough reflection, one described looking for subtle details after giving wrong answers, and two complained about having to answer four yes–no questions for each CMCQ.

Table 7

Student Feedback Regarding Game-Based Study Process

	Learner feedback
Learner1 (traditional group)	Mobile learning makes it possible for me to answer questions anywhere with a cell phone. It is very convenient.
Learner2 (traditional group)	With a leader scoreboard, and also with the feature of badge collections, there are elements of recreation and competition. It gives me some motivation to answer the questions.
Learner1 (confidence group)	I can answer the questions anywhere. It is very convenient. Rewards or penalties are assessed based on my specified degrees of confidence. Therefore, I may get more points if I get it right, but I may also lose more points if I get it wrong. As a result, I will ponder on the correctness of my selections before I submit my answers.
Learner2 (confidence group)	High risk, high pay. This helps me to strengthen my impression of the selections I made in answering the questions.
Learner1 (confirmatory group)	For questions involving concepts that are easy to confuse, reading and answering one option at a time helps me to think in a more structured way. Previously, when I answered (traditional) MCQs, I tended to focus on just one or two options, causing me not to think more thoroughly. Now, when I see an option that I do not know, I will want to understand it even after I have already answered the question.
Learner2 (confirmatory group)	When I answered the questions slowly and at my own pace, it helped me to think more deeply and in a more complete way. Some of the options also helped me to extend my understanding.
Learner3 (confirmatory group)	Becoming used to the idea of reading and answering one option at a time can help me to lessen “reading errors,” “failures of noticing the important points,” and “misunderstandings of what the questions were asking for.”
Learner4 (confirmatory group)	With an appropriate scoring scheme, I will want to find out about the correct answers when I get scores that differed from what I expect. With the goal of getting the scores I deserve, I will want to know the small details of where I get it wrong.
Learner5 (confirmatory group)	I just treated each question as four yes-no questions. It is too much to have to make four selections. I am confident in my answers. Therefore, I always select “Is definitely incorrect” for all other alternatives.
Learner6 (confirmatory group)	To make four selections (in answering each question) is bothersome. The time for answering each question is lengthened to four times as long.

IV. Discussion

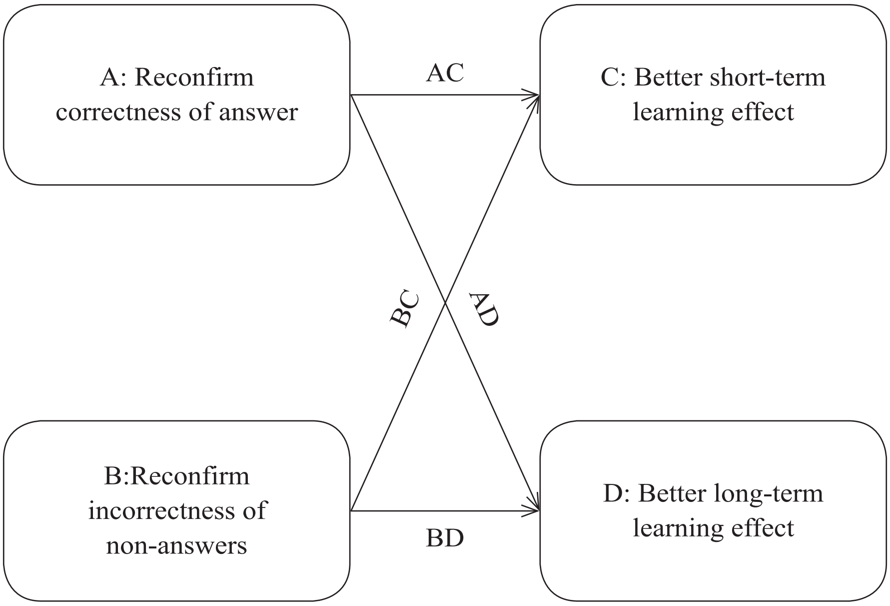
A. Effectiveness of CMCQs for Studying

The CMCQ tests were superior to the MCQ and CBM-based MCQ tests in terms of facilitating study (Tables 1 and 2). At the time of the immediate posttest, the CBM and CMCQ groups had the same understanding of the material ($p > 0.05$, $d = 0.03$). However, the CBM group forgot the material shortly after the experiment, causing them to perform worse than did the CMCQ group on the delayed posttest ($p < 0.05$, $d = -0.49$). Therefore, the CBM-based MCQ tests had a short-term effect on the learning process.

B. Effects of CMCQ Tests on Knowledge Retention

This study performed a causality analysis to determine why the effects of the CBM-based MCQ tests did not last as long as those of the CMCQ tests. Figure 5 depicts four causal relations.

Figure 5
Causality



Node A represents the students checking their answers after a preliminary analysis of the options. Node B represents the students ensuring the other answers were incorrect after a preliminary analysis. Node C represents the superior short-term learning effects. Node D represents the superior long-term learning effects.

Although how A (B) affected C (D) is unknown, three results offer clues. First, a significant difference ($p < .001$) in STRATEGY ratings was observed between the CMCQ and CBM groups, and the difference was large ($d = 1.42$). This suggests that the CMCQ group invested more effort into Node B than did the CBM group. Second, the CMCQ group significantly outperformed the CBM group on the delayed posttest. This suggests that the CMCQ and CBM groups exhibited considerable differences in terms of long-term learning effects (Node D). Third, no significant differences in immediate posttest scores were observed between the CMCQ and CBM groups; this suggests that they did not differ in terms of short-term learning effects (Node C).

During the experiment, both the CMCQ and CBM groups invested some effort into checking their answers (Node A). The ANS-TIME of the CBM group was longer than that of the traditional MCQ group ($d = 0.29$), and the ANS-TIME of the CMCQ group was considerably longer ($d = 0.74$). Therefore, both the CMCQ and CBM groups invested extra time into answering the questions. The extra time the CBM group spent may have been to check their answers, and that of the CMCQ group may have been to check their answers and ensure the distractors were incorrect. The extra time for the CMCQ group was approximately 2.5 times that for the CBM group. Although the CMCQ group checked both the correct and incorrect answers, whereas the CBM group only checked the correct answers, the CBM group invested as much effort into checking the correct answers (Node A) as did the CMCQ group.

The CMCQ and CBM groups differed significantly in terms of UNDERSTANDING ratings ($p < .001$, $d = 1.14$). The CMCQ and CBM groups may not have differed in terms of short-term learning effects (Node C), as suggested by the third result. In this case, the large difference in UNDERSTANDING ratings would support the proposition (based on the second result) that the CMCQ and CBM groups differed considerably in terms of long-term learning effects (Node D).

Therefore, the CMCQ and CBM groups did not differ considerably in terms of checking their answers (Node A) or short-term learning effects (Node C) but differed substantially in terms of checking incorrect answers (Node B) and long-term learning effects (Node D). By investing more effort into checking incorrect answers (Node B), the CMCQ group exhibited superior long-term learning effects (Node D).. Therefore the effort invested in B could determine long-term learning effects. The results of the Pearson's correlation analysis (Table 5) support this conjecture because a moderate correlation (r = .22, $p < 0.05$) was observed between STRATEGY ratings and delayed posttest scores.

C. CMCQ ANS-TIME

The CMCQ group invested more effort into checking both the correct (Node A) and incorrect (Node B) answers than did the traditional MCQ group. As a result, the CMCQ group exhibited superior long-term learning effects (Node D). Some students in the CMCQ group complained about having to answer four questions for each CMCQ. However, the CMCQ ANS-TIME was shorter than expected. The average ANS-TIME of the CMCQ group was longer than that of the traditional MCQ group, but the difference ($d = 0.74 < 0.8$) was not large; subsequent studies should consider informing students that CMCQs require more time.

V. Conclusion

The human intellect consists of the information stored in long-term memory, and "instruction, thus, must consider how [⋯] this information stored and organized in the long-term memory so that it is accessible when and where it is needed" (Lai et al., 2021). The CMCQ-based study technique helped students think comprehensively and thus retain knowledge. In addition, the superior performance of the CMCQ group on the delayed posttest indicates that CMCQ tests should be used as an educational tool.

The grading and analysis of the students' answers to the CMCQs can be automated similarly to that of MCQs. Therefore, CMCQ tests can help students study. In addition, CMCQs require students to reflect and make a subjective, probabilistic estimation for every proposition (Ang & Boo, 2006; Lai et al., 2014). The students who took optional online CMCQ tests retained more than did those who took optional online MCQ and CBM-based MCQ tests, which resulted in superior long-term learning effects.

However, the experimental results indicate that CMCQs required a lengthy ANS-TIME. Therefore, whether CMCQs are suitable for tests requiring high-level cognitive skills such as analysis, evaluation, and creation requires further analysis. Because such questions already require reflection, students may not spend time evaluating the relationships between options when answering. However, the results indicate that CMCQs encourage reflection. Therefore, optional online CMCQ tests offer a viable strategy of helping students learn by enabling them to study at their own pace.

References

Ang, K. C., & Boo, H. K. (2006, November 22-24). *Exploring the use of reflective MCQ in primary science formative assessment* [Paper presentation]. 2006 International Science Education Conference, Singapore. https://repository.nie.edu.sg/bitstream/10497/14904/3/ISEC-2006-86_a.pdf

Candel, C., Vidal-Abarca, E., Cerdán, R., Lippmann, M., & Narciss, S. (2020). Effects of timing of formative feedback in computer-assisted learning environments. *Journal of Computer Assisted Learning, 36*(5), 718-728. https://www.doi.org/10.1111/jcal.12439

Fies, C., & Marshall, J. (2006). Classroom response systems: A review of the literature. *Journal of Science Education and Technology, 15*(1), 101-109. https://doi.org/10.1007/s10956-006-0360-1

Holley, C. D., & Dansereau, D. F. (2014). *Spatial learning strategies: Techniques, applications, and related issues*. Academic Press.

Hsia, Y. T., Jong, B. S., Lin, T. W., & Liao, J. Y. (2019). Designating "hot" items in multiple-choice questions–A strategy for reviewing course materials. *Journal of Computer Assisted Learning, 35*(2), 188-196. https://doi.org/10.1111/jcal.12320

Huang, J., Kerdphol, S., & Inthong, W. (2018). Comparing the effectiveness of rote learning and meaningful learning on English vocabulary memorization for 7th grade students at padoongrasdra school. *Humanities and Social Sciences Journal of Graduate School, Pibulsongkram Rajabhat University, 13*(1), 295-304.

Lai, C. H., Lee, T. P., Jong, B. S., & Hsia Y. T. (2014). Using SPRT+ to Reduce Measure Time on Student Learning Efficiency by Pre-defined Student's Confidence Indicator. *International Journal of Emerging Technologies in Learning, 9*(3), 55-58. https://doi.org/10.3991/ijet.v9i3.3317

Lai, C. H., Jong, B. S., Hsia, Y. T., & Lin, T. W. (2021). Association questions on knowledge retention. *Educational Assessment, Evaluation and Accountability, 33*(2), 375-390. https://doi.org/10.1007/s11092-020-09337-5

Mayer, R. E., Stull, A., DeLeeuw, K., Almeroth, K., Bimber, B., Chun, D., Bulger, M., Campbell, J., Knight, A., & Zhang, H. (2009). Clickers in college classrooms: Fostering learning with questioning methods in large lecture classes. *Contemporary Educational Psychology, 34*(1), 51-57. https://doi.org/10.1016/j.cedpsych.2008.04.002

Novak, J. D. (1990). Concept mapping: A useful tool for science education. *Journal of Research in Science Teaching, 27*(10), 937-949. https://doi.org/10.1002/tea.3660271003

Pamphlett, R., & Farnill, D. (1995). Effect of anxiety on performance in multiple choice examination. *Medical Education, 29*(4), 297-302. https://doi.org/10.1111/j.1365-2923.1995.tb02852.x

Paxton, M. (2000). A linguistic perspective on multiple choice questioning. *Assessment & Evaluation in Higher Education, 25*(2), 109-119. https://doi.org/10.1080/713611429

Ruiz-Primo, M. A., Shavelson, R. J., Li, M., & Schultz, S. E. (2001a). On the validity of cognitive interpretations of scores from alternative concept-mapping techniques. *Educational Assessment, 7*(2), 99-141. https://doi.org/10.1207/S15326977EA0702_2

Ruiz-Primo, M. A., Schultz, S. E., Li, M., & Shavelson, R. J. (2001b). Comparison of the reliability and validity of scores from two concept-mapping techniques. *Journal of Research in Science Teaching: The Official Journal of the National Association for Research in Science Teaching, 38*(2), 260-278. https://doi.org/10.1002/1098-2736 (200102)38:23.0.CO;2-F

Sas, I. C. (2010). *The Multiple-choice concept map (MCCM): An interactive computer-based assessment method*. University of Nevada, Las Vegas.

Schau, C., Mattern, N., Zeilik, M., Teague, K. W., & Weber, R. J. (2001). Select-and-fill-in concept map scores as a measure of students' connected understanding of science. *Educational and Psychological Measurement, 61*(1), 136-158. https://doi.org/10.1177/00131640121971112

Scouller, K. (1998). The influence of assessment method on students' learning approaches: Multiple choice question examination versus assignment essay. *Higher Education, 35*(4), 453-472. https://doi.org/10.1023/A:1003196224280

Snyder, L. G., & Snyder, M. J. (2008). Teaching critical thinking and problem solving skills. *The Journal of Research in Business Education, 50*(2), 90-99.

Tarrant, M., Knierim, A., Hayes, S. K., & Ware, J. (2006). The frequency of item writing flaws in multiple-choice questions used in high stakes nursing assessments. *Nurse Education Today, 26*(8), 662-671. https://doi.org/10.1016/j.nedt.2006.07.006

Tian, Z., Zhang, K., Zhang, T., Dai X. & Lin, J. (2019). Application of Ausubel cognitive assimilation theory in teaching/learning medical biochemistry and molecular biology. *Biochemistry and Molecular Biology Education, 48*(3), 202-219. https://doi.org/10.1002/bmb.21327

Tsai, F. H., Tsai, C. C., & Lin, K. Y. (2015). The evaluation of different gaming modes and feedback types on game-based formative assessment in an online learning environment. *Computers & Education, 81*, 259-269. https://doi.org/10.1016/j.compedu.2014.10.013

2021 年 8 月 17 日收件
2021 年 12 月 11 日通過初審
2021 年 12 月 17 日第一次修正回覆
2022 年 3 月 16 日第二次修正回覆
2022 年 3 月 23 日第三次修正回覆
2022 年 3 月 25 日複審通過