CONSTRUCTIVE MULTIPLE-CHOICE TESTING (CMT) SYSTEM

Jooyong Park

Constructive Multiple-Choice Testing (CMT) System

Jooyong Park
Department of Education, Sejong University, Korea
jpark@sejong.ac.kr

Abstract

A new computerized testing system, called the Constructive Multiple-choice Testing (CMT) system, is introduced. The system combines short-answer and multiple-choice (MC) formats by asking examinees to respond to the same question twice question-by-question, first, in short-answer format, and then in the MC format. An empirical study using the system is reported. Eighty-one 6th grade students (12 years old) were tested on social studies with the CMT system. As expected, students got higher score in the MC format. The correlation between the short-answer and the MC format was medium high (.71). In a subsequent analysis, it was found that whereas those who got intermediate scores in the short-answer part scored relatively higher in the MC part, those who got high or low scores in the short-answer part scored relatively lower in the MC part. Another interesting finding was that there were some cases where students chose a wrong option, even though they typed in the correct answer for the short-answer part. Implications of the results and possible applications of the system are discussed.

Introduction

Item formats are generally dichotomized into the multiple-choice (MC) and the constructed response (CR) formats, with diverse variations in each format (e.g., Bennett, 1993). Ever since the invention of MC tests about 100 years ago, these two formats have been portrayed as antagonistic rivals. Proponents of each format have made biased claims. They have emphasized the advantages of the format that they want to promote, and exaggerated the disadvantages of the other format. Departing from this tradition, I'll propose a new system called the Constructive Multiple-choice Testing (CMT) system. The basic feature of the CMT system is to combine the strengths of the CR and MC formats in a single system. Instead of combining the two formats through separate items as in most current tests, we can combine them at an item level. Examinees are asked to respond to the same stem in two ways, first in the CR format, second in the MC format. Adding the MC format to the CR format restricts the use of the system to short-answer items. Other CR items such as essay and portfolio cannot be accommodated using the system. The immediate benefit of combining the short-answer format and the MC format is that it enables us to collect rich information with no significant

increase in testing time. Other benefits will be discussed later. The present study originated from the effort to overcome the well-known limitations of the MC format. One of them is that examinees are using options as cues to find the correct answer. To solve this problem, we can use innovative computerized technology. In a paper-and-pencil MC test, the options are to be presented along with the stem, whether they are short or extended. However, computer technology allows for alternative methods of presenting the options. One such example can be seen in the Computerized Modified Multiple Choice (CMMT) system (Park, 2005).

The CMMT system was proposed to loosen the tight connection between the stem and the options of an MC item. Loosening the connection between the stem and options can be implemented by presenting stems only and limiting the time to respond to the options. By being given only the stem of a multiple-choice problem, examinees are called upon to generate their answers as if they were solving an open-ended problem. Once the examinee is ready to respond to the question, the examinee can signal to the computer to show the options by clicking the mouse. The options are presented for a short duration: just enough time for the examinee to check his or her answer against the options and choose the right one. Thus, the role of the options changes from cues to matches of answers that examinees have generated. This procedure, as a whole, allows for examinees' active thinking and/or the activation of relevant information from the examinee's memory as in a short-answer test, but permits objective scoring.

In the previous studies, the usefulness of the CMMT as a learning aid has been shown by the enhanced testing effect (Park, 2005; Park & Choi, in press). It can also be used as a delivery system for some innovative item types (Park, 2007). Despite its successful demonstration as a learning tool and as a promising delivery tool, the CMMT system has at least two limitations. One is that random guessing is still possible. The other one is that there is no information about the process. To overcome these two problems of CMMT, the CMT system was proposed.

CMT can be easily characterized in terms of the comparison with the CMMT system. In the CMMT system, stems only are presented first, and students have to respond to the options within a preset time. It is assumed that students would click the mouse for options when they have their own answer in their minds. In the CMT system, instead of having students generate answers in their minds, they are asked to type in the answer on the computer.

In the CMT system, examinees are required to answer the same question twice: first, by typing in their answer, and second, by choosing among the MC options for the same question. Before going into the details, it should be noted that CMT is different from taking the same test twice, i.e., once in the short-answer format and then in the MC format. The first difference is the testing time. It is obvious that CMT will take much less time. More important than testing time, the second difference is that examinees do not have to solve the same problem twice. They can choose the MC option while their CR response is fresh on their minds. It is boring to do the same thing twice. More than that,

when the question is complex and difficult, it will be very stressful for examinees to solve the same problem twice. Thus, practically speaking, it is not possible to have examinees do so for any serious tests.

An empirical study using the CMT system was performed. Eighty-one students participated in the study. They were 6th graders attending P elementary school located in Seoul, Korea. The academic subject tested was 6th grade social studies. The test had 20 questions and covered all the units that the students had learned during the semester. Most of the questions were on factual knowledge. The total testing time was 30 minutes, and the pre-set times for the MC part were set between 4-7 seconds. The main interest was to study the difference between the responses to the two formats.

Results and Discussions

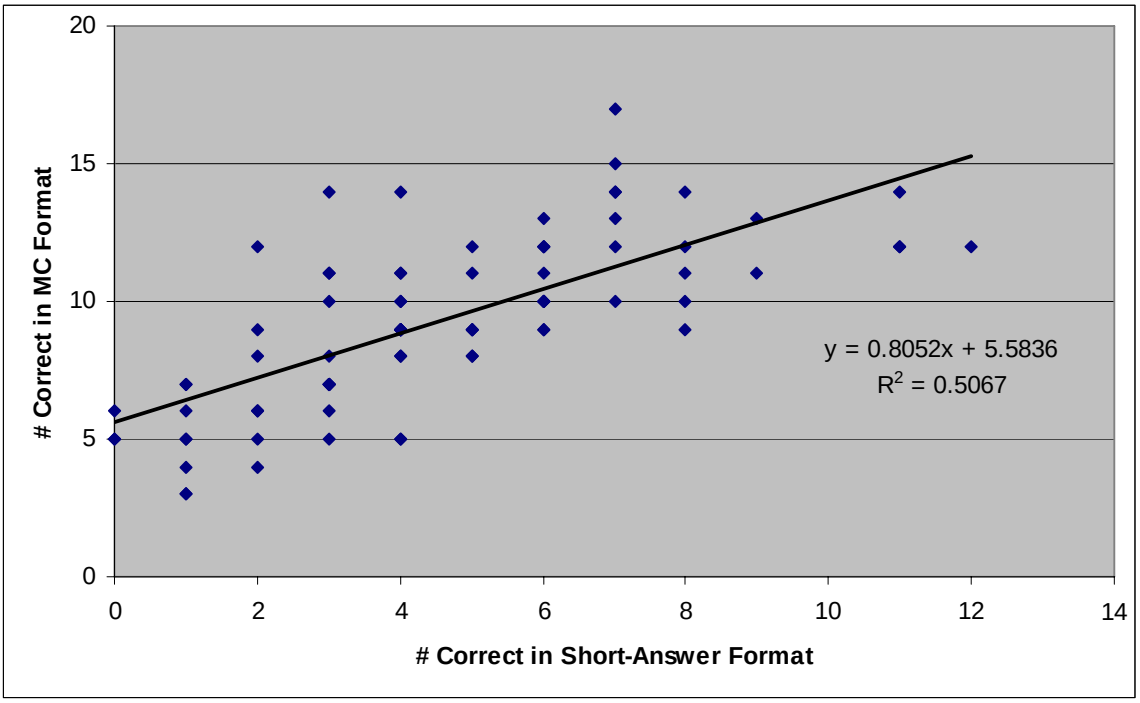
On average, students got higher score for the MC (mean = 9.3, s.d. = 3.1) than for the CR (mean = 4.6, s.d. = 2.3). Correlation between the short-answer format and the MC format was medium high, r = .71. As shown in *Figure 1*, the y-intercept is 5, and the slope is 0.8, showing that the MC part was easier than the CR part. It is also observed that the distribution of the two scores is not uniform along the regression line. Whereas those who got intermediate score in the short-answer part scored relatively higher in the MC part, those who got high and low score in the short-answer part scored relatively lower in the MC part. Part of the reason for this result is that intermediate level students can use the options to find the correct answer even when they cannot generate their own answers. To connect this result with students' level of achievement, students of high and low levels were not affected by the item format, but those of intermediate level got relatively higher score in the MC part. Although more follow-up studies are definitely needed, there is danger that the MC format could mislead students with intermediate level achievement to believe that they have fully understood what they have learned.

Another interesting finding is in the analysis of the MC part for students who got correct answers to the short-answer part. In general, correct answer to the short-answer item is regarded as evidence for solid understanding. Thus, it seems unlikely that students would pick a wrong answer on the MC part after giving the correct answer on the short-answer part. In 11 items, however, there were students who gave correct answers on the short-answer part and yet picked wrong answers on the MC part. Out of the 371 correct responses to the short-answer format, 18 responses (5%) were wrong in the corresponding MC response. It was observed in 11 problems out of 20 in the test. There was a huge variance among them. Seven of them had one wrong answer to the MC format, and 4 items had more than 2 wrong answers to the MC part. Item #3 had 2 wrong answers to the MC out of 7 correct short-answer format. Item #14, #15, and #18 had 2 out of 20, 3 out of 15, and 4 out of 27, respectively. It is possible that the short time limit for the MC part brought about this result. However, if that were the main reason, the error pattern had to be random, rather than systematic as shown. Moreover, for

item #14, 3 out of 4 students picked the same wrong answer for the MC part. It seems that they were misled by visual and phonetic similarities between the correct answer and the wrong option that they ended up picking (correct answer: *yoron*; wrong answer: *unron*). However, the discrepancy shown in the other 3 items cannot be explained by such similarities. This suggests that students were attracted to distractors, even after they had responded correctly to the short-answer format.

It is worth noting that there was a time limit to respond to the MC part. It was set to 4-7 seconds depending on the number of words in options. This was to prevent the students from choosing the right answer by using the information in the options. However, this procedure could have brought about the above result. Thus, the effect of the time limit for the MC part needs to be studied in the near future.

Figure 1. A Scattergram between the Short-answer and the MC scores.



These findings are quite interesting and revealing, but it is too early to draw any firm conclusions. A few future additional studies are listed as follows: The effect of time constraint for the MC part, the effect of allowing the examinees to solve the short-answer format of other items before they go onto the corresponding MC part (in the present study, students had to solve the MC part right after its corresponding short-answer part), the effect of showing the short-answer response when they solve the MC part (in the present study, the examinee's short-answer response disappears from the screen when he or she clicks for the options).

After further empirical studies, CMT system can be applied in following ways. One application is as a unique multiple assessment tool. In recent

assessment literature, many researchers emphasize the use of various assessment techniques (e.g., Wilson & Berenthal, 2006). Considering the emphasis on multiple representation in learning (e.g., Schnotz & Bannert, 2003), the benefits of using diverse assessment formats are undeniable. CMT is an economic and effective tool to implement both MC and CR formats in assessment settings.

Another similar application is to use the CMT as a more accurate testing tool. If examinees receive credit only when they are correct on both the short-answer and the MC, the scoring is more rigorous than any other current scoring method. The students can receive credit only when they are able to generate their own answers and are not misled by attractive distractors. One immediate additional benefit is that teachers' scoring burden is reduced for the short-answer part: Teachers need to grade the short-answer only for correct MC responses.

A third application can include the above two applications, but it is more geared to using tests for learning or formative testing. There is solid evidence that tests enhance learning (e.g., Black & Wiliam, 1998;), and its importance has been emphasized by many researchers (e.g., Earl, 2003; Shepard, 2006). Now using CMT, we can have students involved in the scoring process as follows. Students get feedback immediately after the test based on the MC part. Then they can come to the teacher for partial credit when they are correct on the short-answer part, but wrong on the MC. The short-answer part of CMT can be especially helpful for teachers to trace process information. Although the present CMT program is implemented in a text-only mode, once it is expanded to allow examinees to type in math symbols, we can trace the intermediate products in the subjects of math and science. Teachers can use the intermediate products to find out general pattern of misunderstanding. This can be used for future instruction or generation of attractive distractors in the MC tests. In all, the CMT system can be implemented very easily and can become an effective tool for both test developers and instructors.

Acknowledgements

This work was supported by the Korea Research Foundation Grant funded by the Korean Government (MOEHRD, Basic Research Promotion Fund) (KRF-2007-B00130).

References

Bennett (1993). On the meanings of constructed response. In R.E. Bennett, & W.C. Ward. (Eds.), *Construction versus Choice in cognitive Measurement: Issues in constructed response, performance testing, and portfolio assessment* (pp. 1-27). Hillsdale, NJ: Lawrence Erlbaum Associate, Inc.

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education, 5*(1), 7-73.

Earl, L.M. (2003). *Assessment as learning*. Corwin Press, Thousand Oaks, CA.

Park, J. (2005). Learning in a new computerized testing system. *Journal of Educational Psychology*. *97*(3), 436-443.

Park, J. (2007). A new delivery system for CAT. In D.J. Weiss (Rd.), *Proceedings of the 2007 GMAC Conference on Computerized Adaptive Testing Conference*. Retrieved Aug., 15, 2007 from www.psych.umn.edu/psylabs/CATCentral/

Park, J., & Choi, B. (in press). Higher retention after a New Take-home Computerized Test**.** *British Journal of Educational Technology.*

Schnotz, W. & Bannert, M. (2003). Construction and interference in learning from multiple representation. *Learning & Instruction, 13*, 141-156.

Shepard, L.A. (2006). Classroom assessment. In R. Brennan (Ed.) *Educational measurement* (4th ed.) (pp. 623-646). Preager Publishers.

Wilson, M., & Berenthal, M.W. (2005). *Systems for state science assessment*. National Academy Press.