A New Delivery System for CAT

Jooyong Park

Sejong University, Seoul, Korea

Presented at the Applications and Issues Paper Session, June 7, 2007

2007 GMAC® Conference on Computerized Adaptive Testing

Citation

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

Author Contact

Jooyong Park, mailto:jpark@sejong.ac.kr

A New Delivery System for CAT

CAT has been acknowledged as an effective assessment tool for both higher education admissions tests and major licensure tests. CAT is currently used for the Graduate Management Admission Test (GMAT), for the uniform Certified Public Accountant (CPA) examination, as well as a number of other major testing programs (see the CAT Central Web page at , http://www.psych.umn.edu/psylabs/catcentral/). There is a possibility that the Scholastic Aptitude Test (SAT), American College Testing (ACT) college admissions test, the Law School Admission Test (LSAT), and the Medical College Admission Test (MCAT) will have CAT versions (Zwick, 2006). Such popularity is the result of theoretical and technical advances in many areas of research on CAT (e.g., Darvey & Pitsoniak, 2006; Drasgow, Luecht, & Bennett, 2006), but there is still opportunity for improvement. This paper introduces a new computerized testing system that can be used to overcome some disadvantages of multiple-choice (MC) format and illustrates how the system can be used to deliver some innovative item types.

CAT almost always employs the MC format. In the typical MC format, one salient feature is that ready-made options are given along with the stem. The examinees choose among the options instead of generating their own answers. From this feature comes the main advantage of the MC format, which is automated scoring. Ironically, this feature is also the source of some disadvantages. One disadvantage is guessing. In typical MC tests, examinees can use the options as cues to find the correct answer. By carefully reading the options, examinees can retrieve relevant knowledge, and eliminate implausible options to narrow down the correct answer. In addition to this, presenting options along with the stem has been criticized for its artificiality. That is, the setting is detached from everyday life. One clear example can be found in the title of a research paper on assessing physicians' competence, "Patients don't present with five choices" (Veloski, Rabinowitz, Robeson, & Young, 1999). These shortcomings of the MC format are brought about by the backward reasoning from the options to the answer.

There have been some attempts to prevent backward reasoning. Grid items in SAT math (Braswell & Kupin, 1993), extended matching items, and automatically scorable short-answer items are such examples. The special feature of the grid items is in the answer sheet. Instead of 4 or 5 slots in a row, each item is assigned some columns of 10 or so rows. In the case of 3 columns of 10 rows, when a student gets 9, he or she fills in two 0s in the first two columns and then 9 in the third column. Some mathematical signs can also be included in the answer sheet. One problem is that it is suitable for numerical answers, but not for words.

Some words, such as medical terminology, can be handled in extended matching items (Veloski et al., 1999). The characteristic feature of extended matching items is in the number of its numerous options. Examinees are asked to locate their response on a booklet that contains thousands of options listed alphabetically, along with unique numbers (e.g., heart #3537). Examinees enter the number in their answer sheet. The main problem of this method is the lack of flexibility to deal with other words except for terminologies.

Automatically scorable short answer items have been developed for mathematical items, primarily by Bennett and his colleagues (Bennett, Morley & Quardt, 2000; Bennett, Steffen, Singley, Morley, & Jacquemin, 1997). More recently, c-rater has been developed and used in the National Assessment for Educational Progress (NAEP) and a state-wide assessment in Indiana (Leacock & Chodrow, 2003). However, in order to score short-answer items automatically, a lot of extra work by content experts is needed. In the case of c-rater, for example, for each item a group of expressions with equivalent meaning first have to be identified.

CMMT Items

It is possible to combine the automatic scoring engines with CAT.   However, there is another more economical and flexible way of dealing with the backward reasoning problem. It is using the computerized modified multiple-choice testing (CMMT) system (Park, 2005). The distinct feature of the CMMT system is that it presents the options for a short duration upon the request of the examinee. This manipulation is designed to change the role of the options from cues to matches. As shown in Figure 1a, by being given only the stem of a multiple-choice problem, examinees are called upon to generate their answer as if they were solving an open-ended problem. Once the examinee is ready to respond to the question, he/she can request the computer to show the options by clicking the mouse (Figure 1b). The options are presented for a short duration (5 seconds, in this example) -- just enough time for the examinee to check his/her answer. Once the preset time has elapsed, examinees can no longer respond to the item (Figure 1c). Thus, options are used as matches of answers that examinees generate.

It is obvious that not all MC items can be presented in this new format. Questions requiring the examinee to choose the best answer from among the given options do not fit this method. However, most MC items can be easily converted into the CMMT format. Moreover, the CMMT system can be used to deliver some innovative item types (e.g., Scalise & Gifford, 2006; Sireci & Zenisky, 2006; Zenisky & Sireci, 2002 as illustrated in the following three examples.

The first example is sometimes called "sore-finger" items.   A good example of this item type can be seen in the error recognition part of the Test of English as a Foreign Language (TOEFL).   Examinees are to choose the incorrect word or phrase among underlined expressions:

Guppies are sometimes <u>call</u> rainbow <u>fish</u> <u>because</u> of the males' <u>bright</u> colors.
                       A   B   C   D

To present these items in the CMMT format, the sentence can simply be presented without underlines:
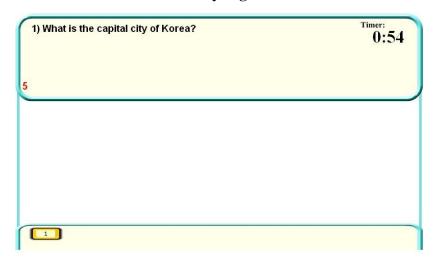
Guppies are sometimes call rainbow fish because of the males' bright colors.

Upon request by an examinee, the correct answer and distractors from the sentence are presented briefly:
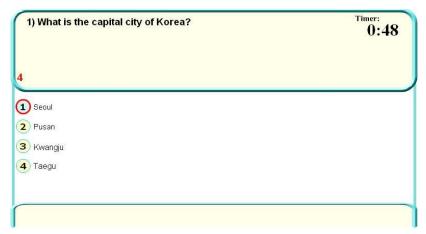
A. call

B. fish

C. because

D. bright

This simple modification increases the item difficulty and decreases the possibility of guessing, because all parts of the sentence become potential options. Some correct sentences can also be included, as well as the none-of-the-above option.
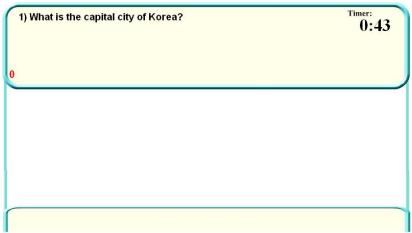
Figure 1. A Sample Display of the CMMT System

a. Stem only is given.

> **1) What is the capital city of Korea?** Timer: **0:54**
>
> 5
>
> 1

b. Options are presented on the request of the examinee.
Examinees can choose one option while the options are on the screen.

> **1) What is the capital city of Korea?** Timer: **0:48**
>
> 4
>
> **(1)** Seoul
> **(2)** Pusan
> **(3)** Kwangju
> **(4)** Taegu

c. When the preset time is over, the options disappear, and the examinee can no longer respond.

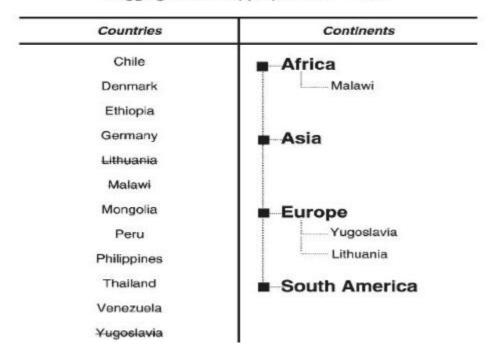> **1) What is the capital city of Korea?** Timer: **0:43**
>
> 0

The second example is create-a-tree items. As shown in Figure 2, examinees are asked to assign the countries (e.g., Chile, Ethiopia, Germany, etc.) in the left-hand column of the display with their respective continents on the right (e.g., Africa, Asia, Europe, etc.).

Figure 2. A Sample Example of Create-a-Tree Items
(From Zenisky & Sireci, 2002, permission requested
from *Applied Measurement in Education*).

Listed below are (a) 12 countries and (b) 4 continents.
Please assign each country to its continent by clicking on the county and dragging it to the appropriate continent.

Countries	Continents
Chile	■ Africa
Denmark	└ Malawi
Ethiopia	
Germany	■ Asia
Lithuania	
Malawi	
Mongolia	■ Europe
Peru	├ Yugoslavia
Philippines	└ Lithuania
Thailand	■ South America
Venezuela	
Yugoslavia	

In converting these items into the CMMT format, embedding the options into the stem is useful, as shown in Figure 3. For example, eight country names (e.g., A. Chile   B. Denmark   C. Ethiopia   D. Germany   E. Malawi   F. Peru   G. Thiland   H. Venezuela) are given along with the question, "Which countries belong to the European Continent?" From the combinations of the eight countries, there are 256 possible answers. Along with the correct answer, which is B, D and G, some subsets of combinations are presented as response options.

The last example is related to graphs and math expressions. In the MC format, examinees can choose the correct answer without knowing how to draw graphs. For instance, examinees can find the right answer by checking whether certain points from each line graph in Figure 4 satisfy the given equation (e.g., $y = 2*x +2$). However, in the CMMT format, because the options are presented briefly for a short time, examinees cannot use such tactics.
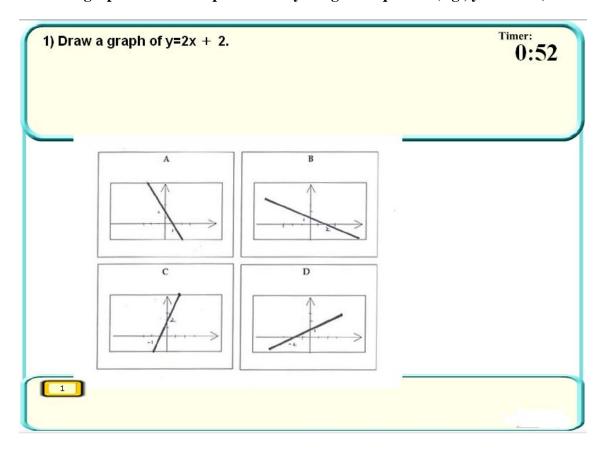
Although these examples are not exhaustive, they are illustrative enough to show the flexibility of the CMMT system. Previous studies have shown that in comparison to the MC format, the CMMT format activates a recall-like process, leading to higher retention than the MC test (Park, 2005; Park & Choi, in press). This has implications for the CMMT system as a learning tool. CMMT format also shows comparable reliability, and increase in item difficulty (Park, in preparation a, b).

Figure 3. A CMMT Version of the Create-a-Tree Item
. Options are embedded into the stem

1) Which of the following countries belong to the European continent?

    A. Chile B. Denmark C. Ethiopia D. Germany
    E. Malawi F. Peru G. Austria H. Mongolia

Timer: **0:51**

7

1

Figure 4. Drawing Graphs in the MC Format
Examinees can find the correct answer by checking whether certain points from each line graph of the four options satisfy the given equation (e.g., y = 2*x +2).

Conclusions

The CMMT system can accommodate other innovative items. Compared to other innovative items types, the CMMT system is a more user-friendly, more flexible, and more economical way of presenting CAT items. As such, it is a promising new delivery system for CAT as well as for computer-based testing (CBT).

References

Bennett, R.E., Morley, M., & Quardt, D. (2000). Three response types for broadening the conception of mathematical problem solving in computerized tests. *Applied Psychological Measurement, 24*(4), 294-309.

Bennett, R. E., Steffen, M., Singley, M. K., Morley, M., & Jacquemin, D. (1997). Evaluating an automatically scorable, open-ended response type for measuring mathematical reasoning in computerized adaptive tests. *Journal of Educational Measurement, 34*, 163-177.

Braswell & Kupin, (1993). Item formats for assessment in mathematics. In R. E. Bennett, & W. C. Ward. (Eds.), *Construction versus choice in cognitive measurement*. New Jersey: Lawrence Erlbaum Associates.

Davey, T., & Pitoniak, M. J. (2006). Designing computerized adaptive tests. In S.M. Downing & T. M. Haladyna (Eds.), *Handbook of test development*. New Jersey: Lawrence Erlbaum Associates.

Drasgow, F., Leucht, R. M., & Bennett, R. E. (2006). Technology and testing. In R. Brennan (Ed.) *Educational measurement* (4th ed.), Preager Publishers.

Leacock, C. & Chodrow, M. (2003). C-rater: Automated scoring of the short-answer questions. *Computers and the Humanities, 37*, 389-405.

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

Park, J. (in preparation-a). *Reliability of the computerized modified multiple-choice testing format*. Department of Education, Sejong University, Seoul.

Park, J. (in preparation-b). *A comparison of item difficulty between the multiple-choice (MC) and the computerized modified multiple-choice testing formats.* Department of Education, Sejong University, Seoul.

Park, J., & Choi, B. (in press). Higher retention after a new take-home computerized test. *British Journal of Educational Technology*.

Scalise, K., & Gifford, B. (2006). Computer-based assessment in E-learning: A framework for constructing "intermediate constraint" questions and tasks for technology platforms. *The Journal of Technology, Learning, and Assessement, 4*(6). Retrieved January 5, 2007 from http://www.jtla.org.

Sireci, S.G., & Zenisky, A.L. (2006). Innovative item formats in computer-based testing: In pursuit of construct representation. In S. M. Downing, & T. M. Haladyna (Eds.), *Handbook of test development,* Lawrence Erlbaum Associates: Hillsdale, NJ.

Veloski, J. J., Rabinowitz, H. K., Robeson, M. R., & Young, P. R. (1999). Patients don't present with five choices: An alternative to multiple-choice tests in assessing physician's competence. *Academic Medicine, 74,* 539-546.

Zenisky, A.L. & Sireci, S. G. (2002). Technological innovations in large-scale assessment. *Applied Measurement in Education, 15*(4), 337-362.

Zwick, R. (2006). Higher education admissions testing. R. Brennan (Ed.) *Educational measurement* (4th ed.), Preager Publishers.