

**A new format for multiple-choice testing: Discrete-Option Multiple-Choice.
Results from early studies¹**

DAVID FOSTER² & HAROLD L. MILLER JR.

Abstract

The standard multiple-choice format has remained relatively unchanged for nearly 100 years, even over the past 25 years as multiple-choice tests have been computerized. We introduce a unique version of the multiple-choice format that has the potential to improve a test's measurement and security properties, along with other advantages. We summarize our research with college students on course-level exams to demonstrate these benefits and to establish the Discrete-Option Multiple-Choice (DOMC) format as not only a viable way to measure skills and content knowledge, but an essential one.

Key words: Discrete-Option Multiple-Choice; multiple choice; test security; fairness; computerized testing

¹ Based on a presentation at the annual Innovations in Testing conference for the Association of Test Publishers, Dallas, Texas, March 3-5, 2008.

² Correspondence concerning this article should be addressed to: David Foster, PhD, 148 S. 1410 E., Lindon, Utah 84042, USA; email: dfoster@kryteriononline.com

Introduction

Traditional multiple-choice test items are ubiquitous in virtually every high-stakes and low-stakes testing program across the world, whether those tests are in the liberal arts and sciences disciplines or in the professions. The multiple-choice format first appeared in the early 1900's when test items were created for use by the United States Army to help select and classify soldiers for military purposes. In its typical form, the multiple-choice format, sometimes referred to as "selected response" format (Downing, 2006), consists of a stem and the answer options, usually 3-5 in number. Of the options, one typically is identified as the correct answer. The others are intended to be incorrect and are called distracters. On a paper exam the test taker marks the answer in some way, or fills in the proper space on a mark-sense answer sheet. Generally, scoring the multiple-choice items is straightforward and objective, and is often accomplished automatically by a computer.

According to Downing (2006) 90 years of research have established the traditional multiple-choice item as an efficient and effective way to measure cognitive achievement or ability. For Downing, most criticisms of the multiple-choice format are not criticisms of the format per se but of poorly developed items.

There is a sizable variety of multiple-choice item types. Many of those that utilize paper-and-pencil presentation have been cataloged by Haladyna (2004). They include the simplest format, often called True/False, and more complex formats, for example, those having a large number of answer options and more than one correct answer. Kubinger and Gottschall (2007) reported a comparative experimental study in which they examined item difficulty as a function of three different formats: six answer options with one correct answer, five answer options with multiple correct answers, and a free-response format.

Computerized versions of the multiple-choice format have extended the variety even further (see Sireci and Zenisky, 2006; Srp, 1994). For example, consider a multiple-choice item with 6 answer options of which 3 are correct answers. The scoring algorithm may require that the test-taker select all 3 in order to answer the item correctly or it may only require any of the 3 to be selected for the item to be scored as correct. These variations on the traditional multiple-choice format, which include manipulating the number and characteristics of the answer options as well as the scoring alternatives, may allow more specific assessment of content knowledge and skills than the more familiar paper-and-pencil versions do.

In computerized versions of multiple-choice items, answer options may be randomized before they are presented to the test taker. This may occur for security purposes – for example, preventing test takers from benefiting by copying from others or having obtained an answer key from someone else. Also, computerized versions may use a scoring algorithm that gives partial credit to the test taker depending on her or his selection of answer options – something that could also occur with the use of paper-and-pencil tests.

Downing (2006) asserts that the multiple-choice format is the "most appropriate item format for measuring cognitive achievement or ability, especially higher order cognitive achievement or cognitive abilities" (p. 288), where "higher order" refers to complex cognitive abilities such as problem solving, reasoning, and analysis and synthesis of information. Support for his statement is easily seen in recent versions of the multiple-choice format that includes the use of computer graphics, animation, audio, video, software simulations, and combinations of these. As each innovation in format has been applied and evaluated, psychometric analyses have remained the standard for judging the effectiveness of the innova-

tion as a contribution to the measurement capability of the assessment instrument that contains it.

Despite the evolution of the multiple-choice format, it remains essentially what it was originally: a stem with answer options that allow the test taker to select from them. We here propose that the evolution of the format take a significant further step, one facilitated by computerization and motivated by the need to improve test security, measurement quality, and fairness. As with previous innovations (see Srp, 1994), we also propose that this new one be evaluated on the empirical evidence that it improves the measurement properties of the assessment instrument as well as its security.

The Discrete-Option Multiple-Choice item format³

The Discrete-Option Multiple-Choice (henceforth DOMC) item format uses the basic elements of the traditional multiple-choice (henceforth Trad-MC) format, namely, the stem and answer options. The essential difference lies in randomly presenting the options one at a time on the screen and asking the test taker to decide if the option that appears is the correct one or not. The item is considered to be completed when the test taker demonstrates that she or he has answered the item correctly or incorrectly. Figure 1 shows an example of a DOMC item using the content of a mathematics question. Note that only one option is presented. In this example, the answer option is the correct answer and was randomly selected for presentation.

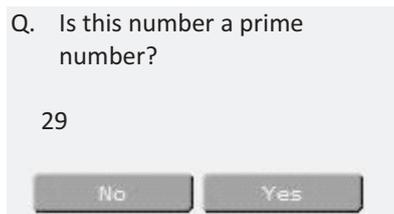


Figure 1:

A typical DOMC item presentation. The DOMC format only displays one answer option on the screen at a time. In this case, "29" was the first randomly selected option

Within the DOMC format, there is only one way for a test taker to answer the item correctly, namely, to choose YES when the correct option is displayed. There are two ways for a test taker to answer a question incorrectly: (a) Choose Yes when an incorrect option (or distractor) is displayed; or (b) choose No when the correct option is displayed. The item continues and provides another answer option if the test taker chooses No when an incorrect option is displayed. Perhaps the best way to understand how the DOMC format works is to experience it directly. Appendix A provides instructions for doing so online.

³ The use of the DOMC is covered by US Patent Number 7,513,775. Licenses to use the DOMC for research purposes are free and available by contacting the first author.

Answer options continue to be presented until the test taker has answered the question correctly or incorrectly, and then the item is scored. With this format, not all of the available answer options need to be presented. The test taker usually answers the item right or wrong before the full set of answer options is exhausted. To reduce feedback about the correctness or incorrectness of a test-taker's response, the DOMC format may incorporate the random presentation of a remaining answer option – which is not scored – with a 0.5 probability *after* the item has been scored. With this feature, test takers are not able to determine reliably when the item will end (unless there are an already known fixed number of options, and all of them have been presented).

Weaknesses of the Trad-MC format

The use of the Trad-MC format has come under intense scrutiny and been the object of criticism as unfair, insecure, and inadequate for the measurement of important skills. Despite Downing's (2006) reassurance that many of the alleged weaknesses are unfounded, two weaknesses of the Trad-MC format are readily apparent. First, it is unfair to many test takers by giving advantage to those who are better at taking such tests or who have cheated to gain an advantage. The former have acquired so-called "test-taking skills" and, being "testwise," have learned how to compare the answer options efficiently, for example, by discovering and exploiting differences between the options and stem of the item. They generally achieve better results than others, who may describe themselves as "poor test-takers" (see Mittring & Rost, 2008). The advantage possessed by those who are testwise is enhanced when items are poorly written, as Downing (2006) asserts, but the advantage nevertheless remains with high-quality test items. According to Crocker, training in testwiseness remains a top concern in the high-stakes areas of admissions testing, licensure, and certification exams, and is becoming more prevalent in K-12 education (Crocker, 2006). Practice with the Trad-MC format can lead to increased scores. Test preparation companies have turned this particular weakness of the format into a multi-billion USD industry, teaching test takers how to gain and improve their test-taking skills *per se* rather than mastering the relevant content that is being assessed.

Another flaw in the Trad-MC format is that items can be easily memorized or captured through technology and shared with others. The entire content of an item is displayed at every instance of test administration, making the content more or less easy to capture and share with others, whether during the test administration or afterward. It is clear to testing programs and test takers alike that Trad-MC items are prone to theft and later re-use. "Brain-dump" sites, that is, Web sites where stolen test content is sold, are proliferating. Test items are often discussed openly on Web forums and in chatrooms. Foster and Zervos (2007) have shown that stolen assessment materials typically are highly accurate, including identification of the correct answer.

In psychometric parlance, test-taking skill and cheating are elements of construct-irrelevant variance (CIV) when they affect the results of multiple-choice assessments. Thus each obtained score represents both relevant knowledge about or skill in the subject matter, plus testwiseness and cheating (and other) components of CIV. Two persons with the same knowledge level or skill level might get very different scores because one is a better test-

taker, a better cheater, or both. As a result, the person *without* the construct-irrelevant skills may be left far behind when scores, grades, certifications, or admissions are awarded.

The DOMC format provides a solution to both weaknesses of the Trad-MC format and thereby may improve psychometric estimation and test security substantially. As for better measurement, it is important to understand the psychometric differences between the Trad-MC and DOMC formats. Logically, the DOMC format may perform better psychometrically by reducing CIV attributable to test-taking skills as well as to cheating. Security is enhanced because not all of the answer options are presented, thereby making it more difficult to memorize or steal items and share that information with others. However, the new format may contribute to CIV in its own way. Because not all of the answer options are presented, or because they are only presented one at a time, this may result in shorter or longer amounts of time to complete each item. We attempted to address these issues, as well as test takers' reactions to the DOMC format, in a series of experiments.

Method

Assessments were administered online to introductory psychology students at Brigham Young University during the 2007-2008 academic year. Each assessment consisted of 20 multiple-choice items delivered in random order online. The items, whether in DOMC or Trad-MC format, each consisted of a stem and 5 answer options. Four of the options were scored as incorrect and were considered distracters. The remaining option was keyed as the correct answer. For both formats, the answer options were also presented in random order. Here we describe three separate experiments. In two of them, Experiment 1 and Experiment 3, we administered the assessments in both Trad-MC and DOMC formats in a counterbalanced design. In Experiment 2 the assessment consisted of DOMC-format items only. Total scores in all three experiments applied to course grading requirements and were provided immediately upon completion of the assessment. None of the students participated in more than one of the experiments.

In Experiment 1, 39 students completed four assessments during the semester. Although the assessments were administered several weeks apart, the data from all four were combined. Each student completed each assessment, which included two 20-item sets with identical multiple-choice content. One set consisted entirely of items with the Trad-MC format; the other contained DOMC-formatted items only. The order of the item sets was randomly established so that half of the students answered the Trad-MC items first and the other half the DOMC items first. To maintain students' motivation for both sets of items, they received the higher of the two scores as part of the course grading requirements. The results of Experiment 1 were used to compare Trad-MC and DOMC items directly in terms of psychometric properties: overall score differences and the item statistics of *p*-values and point-biserial correlations. Item latencies were measured as the time to complete each item from its initial full presentation on the computer screen to the time when the student either submitted it in order to move to the next item when the Trad-MC format was in place or answered it correctly or incorrectly when the DOMC format was in place.

Experiment 2 involved 150 students. Three assessments were included in the experiment and were delivered online. They consisted solely of DOMC-formatted items. The data from the three assessments were combined as in Experiment 1. The larger *n* of Experiment 2

promised more stable item statistics, latencies, and the mean number of answer options that were presented.

Experiment 3 involved 70 students and the addition of survey items. The assessments were identical to those presented in Experiment 1. After the students had completed the first assessment, which contained both DOMC-format and Trad-MC-format items, they were asked to respond to several survey items. Two of these items asked the student to rate how much more difficult or how much easier it would have been to cheat on assessments composed of DOMC-format items compared to those with the Trad-MC format. One of the survey items appears below:

One purpose of the DOMC-question type is to make it more difficult to cheat. Cheating is defined as using some pre-knowledge of the question by getting such information from students who have already taken the test. How much more difficult is it to effectively use such information on the DOMC question compared to the Trad-MC question?

- A. Much more difficult
- B. More difficult
- C. About the same
- D. Less difficult
- E. Much less difficult

Students were also asked about the ability to capture item content by memorization in order to share that content with others who would take the assessment at a later point in time. The survey item appears below:

How much more difficult is it to memorize or copy test-question content so it can be shared with others for the DOMC multiple-choice question compared to the Trad-MC question?

- A. Much more difficult
- B. More difficult
- C. About the same
- D. Less difficult
- E. Much less difficult

We expected the DOMC format to be rated as more difficult in both survey items dealing with test security due to the fact that students cannot easily capture and share the content of such items compared to those with the Trad-MC format.

One other survey item in Experiment 3 addressed the student's preference for one type of format over the other:

Please rate your preference for taking a multiple-choice test using the Trad-MC format compared to the DOMC format.

- A. I STRONGLY prefer the Trad-MC format.
- B. I MODERATELY prefer the Trad-MC format.
- C. I have NO PREFERENCE for either format.
- D. I MODERATELY prefer the DOMC format.
- E. I STRONGLY prefer the DOMC format.

Results

Experiment 1

Total score differences. The results from Experiment 1 showed an 8.59 % drop in the mean total score for assessments using the DOMC format compared to those that used the Trad-MC format. The difference was determined to be statistically significant by a matched t -test [$t = 8.995$; p (one-tailed) = 2.974E-11]. Three of the 39 students had the same score on both assessments, and 2 improved their score on the assessments consisting of items with the DOMC format.

Item analysis. A matched t -test confirmed that, overall, p -values from the assessments using the Trad-MC format were significantly higher than those from their DOMC counterparts [$t = 9.128$; p (one-tailed) = 2.741E-14]. The difference in the mean p -values was 0.107, or 10.7 %. However, a surprising result was that the p -value from the assessments using the DOMC format actually increased for 9 of the 80 items. That is, nine of the DOMC items were actually performing at an easier level than they did in the Trad-MC format. For 8 other items the DOMC and Trad-MC items had the same p -values. Counter to our expectation, the mean point-biserial correlation values were not higher for the assessments using the DOMC format (mean = 0.21) than for those using the Trad-MC format (mean = 0.25). We note that 32 (out of 80) DOMC-format items had higher point-biserial correlations than their Trad-MC counterparts. One item had the same point-biserial correlation with each format.

Item latencies and test times. The average time to complete Trad-MC format items was 24.84 s and 22.68 s for DOMC format items. When extended to 80 items (four assessments), the difference for the mean cumulative testing times was 33.12 min for the Trad-MC format compared to 30.23 min for the DOMC format. This saving of about 10% was statistically significant [$t = 2.799$; p (two-tailed) = 0.0064]. Of the DOMC-format items, 26 had longer latencies than their Trad-MC format correspondents.

Experiment 1 and Experiment 2

Test security. One key to test security is reducing the test-taker's exposure to item content. In the case of the items presented in the Trad-MC format, it is obvious that all 5 answer options for each item were presented simultaneously. This was not the case, for the items that used the DOMC format. A student answering all 20 items in an assessment may see as few as one option for some items or as many as five for others. Each student's experience

with the assessment may be different, based on several factors: the order of items and the answer options that were presented, the difficulty of the items, and the student's level of knowledge of the subject matter. Overall, the students in Experiment 1 saw a mean of 2.64 answer options. For individual items, averaged across students, this value ranged from a low of 1.87 to a high of 3.46. In the replication with larger number of subjects (Experiment 2), the results were similar. The average number of answer options seen by the students was 2.79, and, for individual items, ranged from 2.21 to 3.34.

When we analyzed the mean number of answer options presented across items, we found they were strongly related to the difficulty of the item. That is, the more difficult the item, as judged by its p-value, the fewer answer options were presented. This effect is demonstrated in Figure 2 (Experiment 1) and Figure 3 (Experiment 2). Each of the figures is a scatterplot of the p-value of each item and the mean number of answer options presented for that item.

The correlation coefficients for item p-values and the average number of answer options presented were $r = 0.44$ for Experiment 1 and $r = 0.68$ for Experiment 2. We note that the data contained an additional answer option that was randomly presented for half of the items when a student had technically completed the item, that is, had answered it correctly or incorrectly. The inclusion of this additional answer option was designed to reduce the feedback provided by moving to a new item. Over a 20-item test this resulted in a mean increase of 10 options, thus increasing the mean number of answer options presented. Therefore, the mean number of answer options in the analysis would have been slightly lower had this increment been removed.

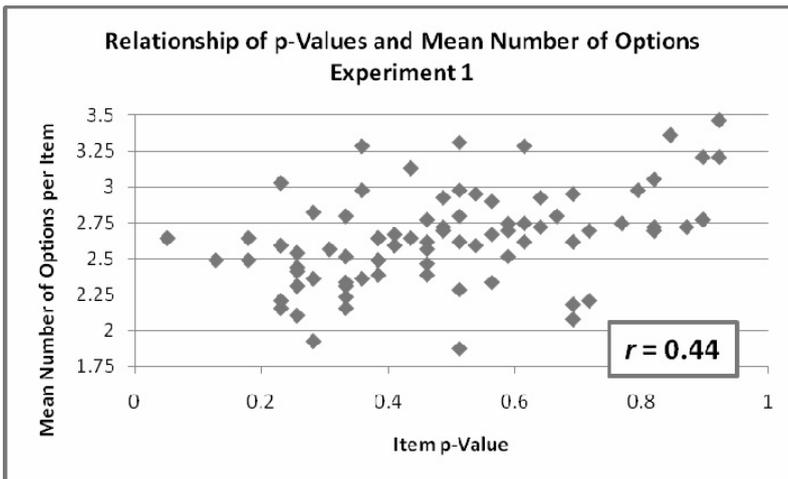


Figure 2:

A scatterplot of item p-values and the average number of answer options for items in Experiment 1

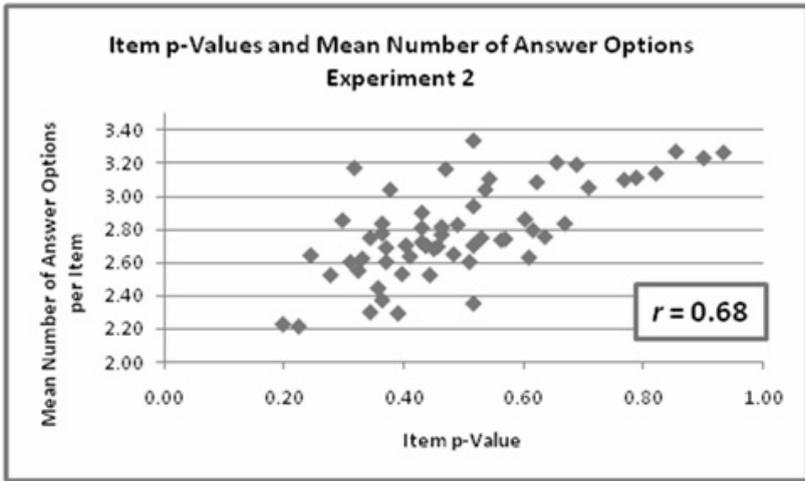


Figure 3:

A scatterplot of item p-values and the average number of answer options for items in Experiment 2

Experiment 3

The results from Experiments 1 and 2 related to test security were further supported by the responses to the two relevant survey items included in Experiment 3 and summarized in Figure 4.

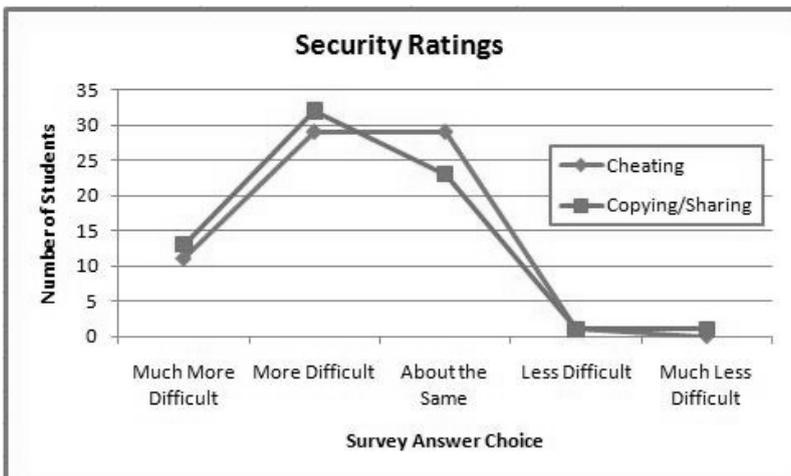


Figure 4:

Survey results for the two test-security survey questions

As Figure 4 indicates, a greater proportion of students judged the DOMC-format items as more difficult to cheat with, memorize, or share with others than items using the Trad-MC format.

Students in Experiment 3 were also asked which of the multiple-choice formats they preferred and how strongly they did so. The results are presented in Figure 5. Fifty-one of the 70 students moderately preferred or strongly preferred the Trad-MC format. Thirteen were indifferent, and only six moderately or strongly preferred the DOMC format.

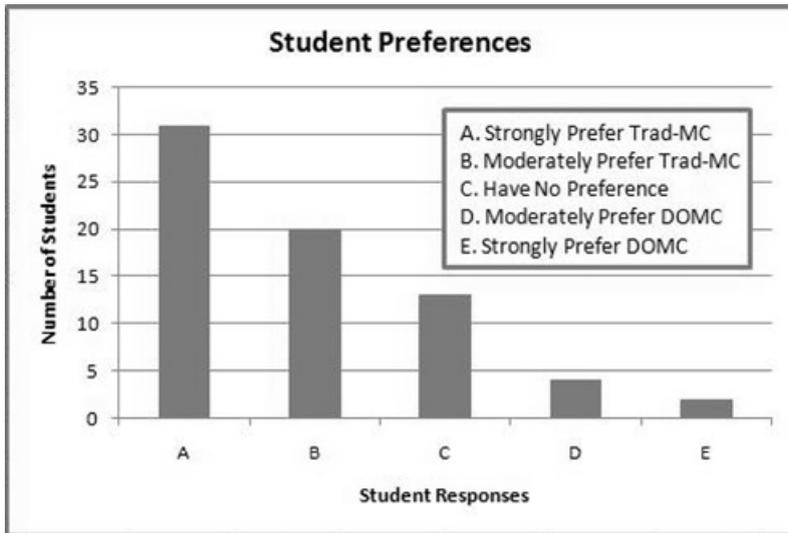


Figure 5:
Survey results for the format-preference question

Discussion

Test scores

There was nearly a 10 % decrease in scores on the assessments composed of DOMC-format items. One possible explanation for this result is that the lower score better represents students' actual level of knowledge of the subject matter. The ability to use test-taking skills or to cheat is significantly reduced, thereby promoting a better estimate of the level of skill or knowledge being measured. As measurement textbooks typically assert, it is critical that test designers understand the potential sources of CIV and design tests to minimize their effects. Although students won't generally appreciate a more difficult assessment format, and initially may be uncomfortable with the DOMC format (see Figure 5 and discussion below), they should appreciate the greater fairness that accompanies an exam that more directly measures their level of knowledge and resists others' efforts to cheat.

Item analysis

Most of the DOMC-format items were more difficult than their Trad-MC counterpart, as evidenced by their lower p-values. We expected this result and account for it in terms of the reduction in CIV. However, for some items there was no difference in difficulty between the formats, or the DOMC-format items were actually easier. What might produce the latter result? One possibility is that there is some aspect of answer options when they are presented in entirety (as in the Trad-MC format) that makes it difficult for the test taker to understand them and answer correctly. It may be that the use of the Trad-MC format offers one or more very attractive distractors that border too closely on being correct. Perhaps the item is a variation on the “choose the *best* answer” type but doesn’t specify such in the stem or in the test instructions. As a *de facto* “choose the *best* answer” type, multiple answer options may be actually correct, even if only partially. Perhaps, when confronted with all the options at once, too many seem correct, and the test taker becomes confused. In contrast, when the item appears in the DOMC format, the correct answer, on average, appears half the time early in the sequence of options. In this way the same opportunity for confusion is not present, and the test taker is more likely to answer correctly. For example, the item shown below may be easier to answer correctly when administered in the DOMC format. While it might not be considered an easy question in either format, the p-value for the item in the DOMC format was 0.28 and for the Trad-MC format, 0.13. Option A is the correct answer.

Q17. Research on hypnosis has shown that much of the behavior that occurs during hypnosis

- A. can also be demonstrated under non-hypnotic conditions.
- B. originates in the unconscious.
- C. is actually a form of imitation.
- D. is dream-like.
- E. is interpreted by the hypnotized subject as useful to her or him.

Table 1 summarizes the responses to the item by the 39 students in Experiment 1. The top half of the table displays the data for the Trad-MC format. Twelve students selected the correct answer; 15 selected one of the distractors, C. The other distractors drew fewer responses. The lower half of the table shows the results for the DOMC format and reveals a different pattern, particularly for option C. Specifically, option C behaved little differently from the other distractors.

Table 1:

The number of students in Experiment 1 who selected each answer option for Q17

Trad-MC	Option A		Option B		Option C		Option D		Option E	
	12		2		15		6		4	
DOMC	Option A		Option B		Option C		Option D		Option E	
Response	Yes	No								
	11	11	7	17	4	17	5	11	3	11

The point-biserial correlation showed that, overall, most items performed similarly regardless of format. However, as with the p-value results, item analysis revealed that some items seemed to have a substantially lower or higher point-biserial correlation when presented in the DOMC format. The item that appears below had a very low point-biserial correlation in the Trad-MC format ($r = 0.05$) and a much higher correlation in the DOMC format ($r = 0.35$). The correct answer is A.

- Q4. Which of the following is an example of negative reinforcement?
- A. Taking an aspirin to get rid of a headache
 - B. Taking away a child’s access to videogames as a consequence of misbehavior
 - C. Patting your dog when he obeys you
 - D. A mother speaking sharply to her naughty child
 - E. Being fined for driving too fast

The results summarized in Table 2 suggest a possible account of the finding.

Table 2:
The number of students in Experiment 1 who selected each answer option for Q4

Trad-MC	Option A		Option B		Option C		Option D		Option E	
	11		13		2		7		6	
DOMC	Option A		Option B		Option C		Option D		Option E	
Response	Yes	No								
	3	17	5	13	0	15	8	10	12	9

Presenting Q4 in the DOMC format turned a difficult question into a very difficult one. However, the 3 students who answered it correctly in the DOMC format were students who had higher total scores on the assessment. When presented in the Trad-MC format, some of the less capable students were able to select the correct answer when they could not in the DOMC format, resulting in the lower point-biserial correlation for that format.

The results shown in Tables 1 and 2 make clear that the two formats interact differently with the content of the question. Problems with multiple-choice answer options may not reveal themselves directly until they are presented in the DOMC format. This may make it useful for item developers as they attempt to produce high-quality, unambiguous items (see Moreno, Martinez, & Muniz, 2006). Using the DOMC item for multiple-choice items that provide answer options that are *all* correct and which require the test taker to choose the BEST answer, is not advised, as the premise of the item design requires the comparison of answer options. Fortunately there is no *a priori* psychometric restriction to mixing the DOMC and Trad-MC together on the same exam.

Item latency differences

Because the assessments were computerized, it was possible to accurately measure differences in the time to complete individual items. Overall, the items using the DOMC format were answered in about 10% less time than those using the Trad-MC format. Having to consider fewer answer options overall may reduce the time it requires to answer a specific item.

In related research (Foster, unpublished research) with very young English-language learners, ages 3 to 6, the DOMC format replaced a Trad-MC format that had 3 answer options. Stems and answer options for assessments measuring the English competency of the children were all comprised of video, animation, graphics, or audio. Text was used only for letter-recognition items. Those children with a tentative grasp of the English language (according to their parents and pre-school or kindergarten teachers) struggled with the assessment that contained items using the Trad-MC format, usually taking a very long time to complete it, if at all. With the DOMC format, however, the children generally completed the assessment and did so in much less time. It may be argued that the Trad-MC presented the children with a very difficult cognitive task in which they had to understand the stem and then compare the answer options, all in a new computerized testing format and a relatively unfamiliar language. The task was made simpler with the DOMC format. Provided the child understood the stem, she or he could select the correct answer option on the basis of what she or he knew about the language rather than relative to the other answer options.

Test security

Were the DOMC format to deter or prevent cheating to a greater degree than the Trad-MC format, it would be seen as a significant advance in the testing industry's battle against increasingly successful attempts to cheat and steal test content (Cizek, 1999; Cohen and Wollack, 2006). The survey results from Experiment 3 showed that students considered the DOMC format much more difficult to cheat on and steal from. This may be attributed to the fact that fewer answer options were presented with the DOMC format. If a test taker wished to share what he or she saw in the test with another test taker at a later time, the description likely would lack some of the answer options per item. Moreover, there is no guarantee that the description would include the correct answer. The person receiving the description would be working from partial information at best.

We found that fewer answer options are displayed for more difficult items than for easier ones. A possible explanation for this unexpected finding is that more difficult items tend to be resolved earlier than easier ones. That is, students tended to answer more difficult questions incorrectly with fewer answer options presented. Difficult questions are generally more discriminating and may also be more expensive to produce. The DOMC format provides a convenient way of protecting them differentially.

According to Bennett (1999), the use of new computer technology that automatically generates item variants from a parent item will enhance test security. The DOMC format generates different variants of the same item each time the assessment is presented, thus increasing the difficulty of capturing and sharing test content.

General discussion

In this paper we have described a unique multiple-choice testing format and reported initial research that compared it to the conventional format. The effects were seen in the increased difficulty and discriminative ability of items, reduction of the time needed to complete the assessment, and the improved security of the assessment. Moreover, these effects come with minimal burden to item writers, as most items can be revised readily to accommodate the new format.

One other effect is worth noting and may deserve a research program all its own. Because it reduces the effects of test-taking skills as well as attempts at testing fraud, the DOMC format improves the fairness of the assessment. Those who do not cheat, and those who cannot afford often expensive test-preparation courses, now can take exams on a more level playing field, as it were.

Messick (1989) argues that validity, an important quality of any test, involves understanding the consequences of a test or, more accurately, the consequences of producing test scores using that test. As a relevant example, Linn, Baker, and Dunbar (1991) considered the intended and unintended consequences of moving to performance-based testing formats for educational exams. They asked whether and how the results of these exams might change the way teachers teach their students? Furthermore, would the consequences of those changes be positive or negative? Similarly, there are known, unintended negative consequences of continuing to use Trad-MC format, such as teaching to the test and over-reliance on test-taking skills. We fully expect that use of the DOMC in small- or large-scale assessment programs will lead to renewed activities, including in-class activities, that focus on learning and understanding the content associated with the exam. Unintended consequences are expected as well. The preliminary reactions of students to the DOMC format that we report here are mixed, not least because the new format obviates the advantage of the test-taking skills on which many students have relied. It obliges them to take a different, possibly more demanding approach to studying and preparing for exams.

For psychometric, security, and fairness reasons, it is important to reconsider the ubiquitous reliance on the traditional multiple-choice format. With computers increasingly available for administering assessments, it is increasingly feasible to convert to the “new and improved” multiple-choice, the one that offers discrete options.

References

- Bennett, R. E. (1999). Using new technology to improve assessment. *Educational Measurement: Issues and Practice*, Fall, 1999, 5-12.
- Cizek, G. (1999). *Cheating on tests: How to do it, detect it and prevent it*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Cohen, A. S., & Wollack, J. A. (2006). Test administration, security, scoring, and reporting. In R. L. Brennan (Ed.), *Educational Measurement*, 4th ed. (pp. 355-386). Westport, CT: American Council on Education/Praeger.
- Crocker, L. (2006). Preparing examinees for test taking: Guidelines for test developers and test users. In S. M. Downing & T. M. Haladyna (Eds.), *Handbook of Test Development* (pp. 115-128). Mahwah, NJ: Lawrence Erlbaum Associates.

- Downing, S. M. (2006). Twelve steps for effective test development. In S. M. Downing & T. M. Haladyna (Eds.), *Handbook of Test Development* (pp. 3-38). Mahwah, NJ: Lawrence Erlbaum Associates.
- Foster, D. F., & Zervos, C. (2007). *The big Internet heist*. Poster presented at the annual meeting for the Association of Test Publishers, Orlando, FL.
- Haladyna, T. M. (2004). *Developing and validating multiple-choice test items*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Kubinger, K. D., & Gottschall, C. H. (2007). Item difficulty of multiple choice tests dependent on different item response formats – An experiment in fundamental research on psychological assessment. *Psychology Science Quarterly*, 49, 361-374.
- Linn, R. L., Baker, E. L., & Dunbar, S. B. (1991). Complex, performance-based assessment: Expectations and validation criteria. *Educational Researcher*, November, 15-21.
- Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational Measurement*, 3rd ed. (pp. 13-104). New York: Macmillan.
- Mittring, G., & Rost, D. H. (2008). Die verflixten Distraktoren – Über den Nutzen einer theoretischen Distraktorenanalyse bei Matrizentests. *Diagnostica*, 54, 193-201.
- Moreno, R., Martinez, R. J., & Muniz, J. (2006). New guidelines for developing multiple-choice items. *Methodology*, 2, 65-72.
- Sireci, S. G., & Zenisky, A. L. (2006). Innovative item formats in computer-based testing: In pursuit of improved construct representation. In S. M. Downing & T. S. Haladyna (Eds.), *Handbook of Test Development* (pp. 329-348). Mahwah, NJ: Lawrence Erlbaum Associates.
- Srp, G. (1994). *Syllogismen* [Software and Manual]. Frankfurt/M.: Swets Test Services.

Appendix A

Instructions for taking sample tests in the DOMC item format

The best way to understand how the DOMC format works is to experience it. The reader can take a sample test that uses items with the DOMC format by following these steps:

1. Go to www.webassessor.com.
2. Enter Login: fosteritem.
3. Enter Password: samples.
4. Select: Register for a new assessment.
5. Select: Get Now for the test you want to take.
6. Select: Checkout.
7. Select: Done.
8. Find the scheduled test you chose from the list of tests, then Select: Launch.

Note that the items included in the sample tests have not been optimized for the DOMC format.