Scholarship of Teaching and Learning in Psychology
2016, Vol. 2, No. 2, 147–158

2332-2101/16/$12.00 http://dx.doi.org/10.1037/stl0000062

TEACHER-READY RESEARCH REVIEW

Multiple-Choice Questions: Tips for Optimizing Assessment In-Seat and Online

Xiaomeng Xu, Sierra Kauer, and Samantha Tupy
Idaho State University

Multiple-choice questions are frequently used in college classrooms as part of student assessment. While multiple-choice assessments (compared to other formats such as constructed response) seem to be the preferred method of testing by instructors and students, their effectiveness in assessing comprehension in college courses is sometimes called into question. Research has shown that there are ways to optimize the construction and use of multiple-choice testing to benefit college classroom instruction and assessment, student learning, and performance, and to more efficiently utilize instructor's time and energy. This teacher-ready research review provides an overview of the research on utilizing multiple-choice questions as well as some tips on using, writing, and administering multiple-choice questions during assessments. We also summarize the benefits and potential issues with using multiple-choice questions including concerns about cheating, ways to detect and deter cheating, and testing issues and strategies unique to online formats. We hope that this short review will be helpful to instructors as they prepare courses and assessments and further encourage the use of empirical data in pedagogy related decision-making.

Keywords: multiple-choice, testing, exams, assessment, pedagogy

Multiple-choice questions are commonly used for assessment in college classrooms. Introductory and lower-level courses are especially likely to feature quizzes and exams that contain at least some multiple-choice items as class sizes tend to be larger. Efficiency in grading tests is also one reason multiple-choice is utilized in standardized testing for placement into college and graduate school. Research indicates that instructors may have a preference for multiple-choice exams not only because it makes administering and grading the exam simpler, but because it allows for objective and consistent grading (Simkin & Kuechler, 2005; Zeidner, 1987). Students also tend to prefer multiple-choice questions due to reduced instructor bias and because they consider multiple-choice question exams to be easier since they can choose answers based on the process of elimination (Simkin & Kuechler, 2005; Struyven, Dochy, & Janssens, 2005; Tozoglu, Tozoglu, Gurses, & Dogar, 2004; Zeidner, 1987). This preference by instructors and students for multiple-choice exams has led to high use of multiple-choice testing in college. Despite the preference by many instructors and students for this format, multiple-choice question assessments are sometimes called into question regarding their ability to assess comprehension (Ozuru, Briner, Kurby, & McNamara, 2013).

In this teacher-ready research review, we summarize and discuss the literature on multiple-choice testing (focusing on assessment, feedback, and efficiency), highlight valid concerns, and provide tips for attenuating or bypassing potential issues (see Table 1 for some highlights). These tips are by no means exhaustive, and other reviews of multiple-choice testing and strategies exist, perhaps most notably Haladyna, Downing, and Rodriguez's (2002) taxonomy of 31 multiple-choice item-writing

Xiaomeng Xu, Sierra Kauer, and Samantha Tupy, Department of Psychology, Idaho State University.

Correspondence concerning this article should be addressed to Xiaomeng Xu, Department of Psychology, Idaho State University, 921 S 8th Avenue, Stop 8112, Pocatello, ID 83209. E-mail: xuxiao@isu.edu

Table 1
Optimizing Multiple-Choice Question Testing

Domain	Optimization techniques
Assessment quality	Utilize questions designed for higher-order cognitive assessment.
	Discourage students from guessing and/or utilize methods of scoring that penalize guessing.
	Improve quality of exams by conducting item analyses.
	Utilize collaborative testing when appropriate.
Fairness	Use questions that are clearly written.
	Write questions that cover a broad range of topics.
	Use questions that are consistent with the syllabus.
	Inform students about the knowledge to be assessed.
Feedback	Utilize elaborative feedback.
	Timing (i.e., immediate or delayed) of feedback should be based on difficulty level of item and context of the assessment.
	Give students opportunities to self-correct.
	Provide elaborate and timely feedback for online assessments using software.
	Solicit feedback on assessments from students.
Formatting and content	Utilize 3-choice items.
	Avoid questions that use negatives.
	Avoid multipart and giveaway questions.
	Avoid "none of the above" questions.
	Avoid composite answers such as "A and B but not C."
	Avoid "all of the above" questions.
	Choices should be parallel in structure and equal in length.
	Question stems should be as short as possible, but contain all relevant information.
	Randomize answer positions.
Cheating countermeasures	Use alternate test forms.
	Utilize alternate or assigned seating for assessments.
	Provide students with an academic integrity policy.
	Utilize honor codes and academic honesty agreements.
	Draw from a large question bank and randomize question answers and order.
	Defer question feedback until after online assessments close.
	Change exam questions between semesters.
	Utilize "lockdown browsers" for online classes.

guidelines. This review outlines some of their key points but also provides further information based on more recent empirical literature. In addition, this review also addresses multiple-choice testing and optimization specifically in the context of online assessments (e.g., in online or hybrid courses).

Assessment

Effectiveness and Assessment Quality

Constructed-response exams consist of questions where the test-taker needs to produce an answer rather than selecting from provided options (as per multiple-choice exams). These constructed-responses products may be short such as with fill-in-the-blank or longer such as with short answers and essays. Compared to constructed-response exams, multiple-choice exams can be effective, yielding similar test scores (e.g., Hickson, Reed, & Sander, 2012) or even higher test scores (e.g., Park, 2010). Multiple-choice testing has been shown to positively enhance retention of the material that is tested (a testing effect) and to boost performance on later tests (see Marsh, Roediger, Bjork, & Bjork, 2007, for a review). Students also report that multiple-choice questions bolster confidence and self-esteem and are useful for learning basic concepts (Douglas, Wilson, & Ennis, 2012). It is important to note however that one reason students may typically prefer multiple-choice questions is because they perform better on them than on short-answer questions, which may provide misinformation to instructors about true student learning (e.g., Funk & Dickson, 2011).

Multiple-choice questions may improve quality of testing as it enables test-givers to ask a greater number of questions on a

broader set of topics (which can contribute to the reliability of the assessment), increases perceived objectivity in the grading process (e.g., ensuring that students do not lose points for poor spelling, grammar, or writing skills), reduces subjectivity/inconsistency/human errors in scoring (particularly when machine graded), and can reduce student anxiety (Simkin & Kuechler, 2005). Multiple-choice testing also allows instructors to calculate exam statistics such as item-total correlations (sometimes referred to as discrimination coefficients) which are the correlations between scores on a particular item (dichotomized into 0 for incorrect and 1 for correct) and overall test scores (see Davis, 1993, for instructions on how to calculate discrimination levels by hand if correlation coefficient computer programs are not available), and utilize these statistics in improving the quality of subsequent exams.

Shallower Assessment

Instructors may be concerned that multiple-choice questions do not assess comprehension, higher-order cognitions, or deeper understanding of the topic the way short-answer and essay questions can (e.g., Ozuru et al., 2013). It is also more difficult to pinpoint student's true knowledge (e.g., why they got the question right/wrong, if they were guessing) and to provide partial credit with multiple-choice testing. Related to this issue, students are also more likely to perceive multiple-choice tests as assessing lower-level cognitive processing (e.g., memorizing facts) and thus use more surface-learning approaches when they are aware that multiple-choice testing will be used (e.g., Scouller, 1998; Scouller & Prosser, 1994; Yonker, 2011). Lower-level courses with large enrollments such as introductory psychology classes are more likely to utilize (sometimes exclusively) multiple-choice testing. Introductory psychology students also exhibit low levels of course material retention (e.g., 56% accuracy 2 years after the class), and it is possible that multiple-choice tests, as well as the way students perceive these tests and shallowly study for them (focusing on recall rather than meaning and understanding), may contribute to this issue (Landrum & Gurung, 2013; Scouller, 1998; Scouller & Prosser, 1994).

Assessing Deeper Thinking

While many instructors utilize multiple-choice questions (effectively) to assess more superficial or rote-learning understanding (e.g., definitions, identifying a lobe of the brain), it is also possible for multiple-choice questions to assess higher-order cognitive skills, synthesis, application, and other components of deeper understanding and thinking (e.g., Simkin & Keuchler, 2005; Tractenberg, Gushta, Mulroney, & Weissinger, 2013).

To assess more higher-order cognitive skills, test-makers can create multiple-choice questions that ask for the best answer available (e.g., what was the main reason, which is most likely), with multiple choices offering plausible (even good) answers but in which only one is the best answer. Deeper-thinking questions may also ask students to identify the theory that an example illustrates or to predict the outcome of a hypothetical (Davis, 1993). Some examples of multiple-choice questions that assess deeper understanding and thinking have been shared by psychologists such as Steven Pinker (Steven Pinker's mind games, 2014), and via helpful assessment tools by teachers/researchers in other fields (e.g., the Blooming Biology Tool; Crowe, Dirks, & Wenderoth, 2008).

Different formats of multiple-choice questions can also be used to assist in assessing deeper thinking and understanding. One example is the ordered multiple-choice testing format (Briggs, Alonzo, Schwab, & Wilson, 2006), where each answer choice represents a different developmental level of understanding. Thus interpreting item responses provides instructors with a better grasp of how deeply a student understands the content.

Another method to try to ensure that students' true understanding of the topic is assessed (rather than random guessing) is to conduct confidence testing (see Davies, 2002), where students are required to report level of confidence in knowing the answer prior to selecting (or even being able to view with online testing) answer choices. Scoring is then based on a combination of whether or not students selected the correct answer, as well as how confident they were in their response. Thus, a confident

response paired with a correct answer would score higher than a nonconfident response (e.g., guessing) paired with a correct answer. This method may be problematic, however, due to logistics (e.g., increased grading time if calculation of augmented scores is not automated) and concerns about individual differences (rather than actual knowledge) affecting how confidently students respond. Some research has shown that personality affects confidence reporting and scores (R. Hansen, 1971; Jacobs, 1971), although other research has shown that correlations between confidence testing scores and personality do not persist once students have had practice with confidence testing and their performance is taken into account (Echternacht, Boldt, & Sellman, 1972).

Foster and Miller (2009) provided another format (discrete-option multiple choice) that may better assess true understanding than typical multiple-choice formats. This format (optimized for online testing) randomly presents the answer options for the question one at a time to the student who selects whether or not that option is correct. Jensen et al. (2006) also described another format ("You are the teacher" questions), which asks students to imagine that they are a teacher correcting a short-answer/essay and their goal is to identify all errors in the passage they are reading. The multiple-choice question therefore is a short paragraph and the answer choices are different numbers of errors (e.g., 0, 1, 2, 3, 4, or more).

Finally, instructors concerned about guessing may explicitly discourage students from guessing or utilize a method of scoring that does so (e.g., penalizing for incorrect answers as per some standardized tests such as the SAT). One method is formula scoring which occurs when a proportion of the number of incorrect response is subtracted from the number of correct responses (see Frary, 1988).

Low-Quality Questions and Answers

Multiple-choice questions can be effective, but there is room for improvement on their quality. For example, DiBattista and Kurzawa (2011) analyzed 1,198 multiple-choice items from exams across 16 undergraduate classes and found that many were flawed. The most common problems included nonoptimal incorrect answers (i.e., selected by fewer than 5% of examinees) and unsatisfactory item-total correlations (i.e., less than +.20). Based on the results of their analyses, DiBattista and Kurzawa (2011) suggested that test-makers improve the quality of multiple-choice exams by conducting an item analysis after administering a test and removing or replacing incorrect items that weaken discriminatory power. DiBattista and Kurzawa also noted that postsecondary institutions should provide training for faculty on testing (including how to interpret an item analysis report) and support for faculty to improve their assessments (including providing faculty with an item analysis report following every multiple-choice exam). Instructors may want to pay special attention to common errors of students (on assessments as well as in class) as these can be utilized as high-quality incorrect answers (see freely available workshop material by Zimmaro, 2010, for this and additional tips on writing good multiple-choice exams).

Fairness

McCoubrie (2004) reviewed the literature on multiple-choice questions and offered suggestions on how to improve the fairness of these questions. Fairness is an important aspect for test-makers to keep in mind, as students who perceive assessments to be fair are more likely to study and learn the material rather than studying just for the test (particularly if the exam is high-stakes). Important factors that contribute to fairness include ensuring that items are clear (e.g., questions and answers are not confusing), that they cover a broad range of topics, and are consistent with the syllabus (e.g., not testing only a small proportion of what is discussed in the course). Also, student perceptions of testing influence how they choose to study and multiple-choice questions tend to be seen as requiring lower-level cognitive processing (Scouller, 1998). Therefore, it may be beneficial to provide students with information about multiple-choice exams to assist them in preparing appropriately (e.g., letting students know if multiple-choice questions will be assessing only factual information or will require deeper understanding and critical thinking skills).

False Knowledge and Memory

How students study for multiple-choice question exams can influence memory of concepts and answers to questions. For example, intro-

ductory psychology students have been shown to rate answers as true based on familiarity of the question item (Begg, Armour, & Kerr, 1985). Thus, mere familiarity of an incorrect item can influence choice of test question answers. Furthermore, students may be exposed to false information in the form of incorrect answers to multiple-choice questions and may remember this misinformation as true at later times. Therefore, multiple-choice questions may lead to false knowledge (Roediger & Marsh, 2005; Marsh et al., 2007). However, there are ways to attenuate this negative testing effect. One way to enhance student learning and help with potential false memory issues is to use collaborative testing where student work in pairs or groups. Collaborative testing has been shown to lead to better performance on assessments (Rao, Collins, & DiCarlo, 2002) as well as better retention of knowledge 4 weeks later (Cortright, Collins, Rodenbaugh, & DiCarlo, 2003). While collaborative testing can reduce anxiety, the mechanisms through which exam scores improve are having good discussions, remembering information better, and an increased ability to think about the tested information (Kapitanoff, 2009). Another way to attenuate the false memory effect is by providing timely and appropriate feedback (Butler & Roediger, 2008).

Feedback

Content of Feedback

Consistent meaningful feedback (e.g., detailed explanations of why certain answers were correct or incorrect) is an important component of student learning outcomes, enjoyment, engagement in the course, and ratings of teaching quality (Gaytan & McEwen, 2007). Elaborated feedback has been shown to be more effective for learning when compared to simple answer verification (Pridemore & Klein, 1991). Additionally, formative feedback (aimed at modifying thinking or behavior to improve learning) needs to also be supportive and nonevaluative (e.g., judging the response, not the student; see Shute, 2008, for guidelines on providing formative feedback). More elaborate and detailed feedback also tends to require more effort on the part of the instructor, which may (in part) explain why not all students who take multiple-choice assessments receive this type of feedback. However, there are strategies that can be utilized (including technology) which reduce instructor burden when providing high-quality feedback (see Feedback Techniques section, below).

Timing of Feedback

An important component for student learning is to provide timely feedback after testing (e.g., Butler, Karpicke, & Roediger, 2007). Multiple-choice exams facilitate this process as these exams tend to be quicker and less subjective to grade (particularly as grading can be computerized). Many recommend that feedback be provided as soon as possible after testing (e.g., Gaytan & McEwen, 2007), although some research indicates that a short delay (e.g., 10 min) is optimal (produces better outcomes than immediate feedback) as it spaces out the learning (Butler et al., 2007). Hattie and Timperley (2007) reviewed the literature on feedback and found that across eight meta-analyses (398 studies), immediate feedback is generally superior to delayed feedback (see their Table 2). However, they pointed out that there are additional important considerations when determining timing. For example, delayed feedback becomes more effective (and superior to immediate feedback) as the difficulty level of the item increases (e.g., Clariana, Wagner, & Roher Murphy, 2000) potentially because the delay allows for greater processing. Similarly, the effects of timing depend on the context of the assessment with immediate feedback being more effective in studies that use actual quizzes and delayed feedback being more effective in laboratory studies that use list learning (Kulik & Kulik, 1988).

Feedback Techniques

One method of providing higher-quality feedback while minimizing instructor burden is to offer students opportunities to self-correct. Grühn and Cheng (2014) found that students who were allowed to hand in a self-corrected midterm performed better on the final exam compared to students who took a traditional midterm (and were not allowed to self-correct). Grühn and Cheng utilized procedures for self-correction as laid out by Montepare (2005, 2007). Specifically, students take an exam as

per usual and hand in their (original) answers and are then provided with a copy of the questions to take home. Students then hand in a second set of answers (self-corrected) after they have had time to review the material. Scoring takes both the original and self-corrected responses into account (full credit for questions that are correct on both sets of responses, partial credit for questions that were initially incorrect but correct in the revised set of responses, no credit for questions that are incorrect on both sets of responses). Thus, the self-correction method focuses on fostering additional learning and rewards students' additional time and energy spent with the course material. Learning can also be facilitated with instructor- or computer-graded assessments if students are allowed to resubmit based on feedback (e.g., Cole & Todd, 2003).

For online multiple-choice testing, detailed feedback (immediate or delayed) can be automatically provided to students based on the correct (or incorrect) items that they choose (instructors can also program the settings such that students can view all feedback for all answer choices, regardless of what they got correct or incorrect). While this feedback needs to be programmed into the online system, once it has been input it can be provided automatically to all students without the need for the instructor to individually mark up each exam, write the same comments over and over, or go over each exam question and answer in class.

Instructors providing feedback need to decide how to give students access to the questions and their answers—that is, whether or not students' exams are returned to them. In courses where large high-quality test banks are available, instructors may have the ability to utilize new questions each semester and thus may return exams without worrying about students copying questions and answers. However, many instructors recycle at least some of their questions and need to balance optimal feedback (e.g., immediate, elaborate) with pragmatic concerns (thus techniques such as self-correcting may not be as appealing for some instructors). In face-to-face settings, instructors can supervise review of the exam and return exams temporarily (e.g., for a portion of a class period or during office hours). For online settings, instructors can program feedback to be temporary and delayed until after the exam closes for everyone (thus students cannot see feedback while the exam is still "live" for other students, preventing the sharing of answers). In some situations, these efforts may not be enough to deter academic dishonesty and instructors may need to utilize other techniques (see Cheating Prevention and Countermeasures section, below).

Finally, in addition to providing students with feedback on their responses and performance, it may be useful to solicit feedback from students, allowing them an opportunity to comment on the exam and for instructors to utilize this data in future assessments (Davis, 1993).

Efficiency

Test Administration

Simkin and Kuechler (2005) presented a review of the literature on multiple-choice questions compared to constructed response questions. Some benefits of multiple-choice include efficiency in testing large numbers of students and grading large numbers of exams (e.g., scoring via computerized mechanisms), which also facilitates providing students with timely feedback. Additionally, multiple versions of the same multiple-choice exam are relatively easy to administer (which may help deter cheating) and test items can be stored, edited, and reused fairly easily, helping to reduce the time needed for test-makers to design exams (this is reduced even further if instructors have access to test banks).

Format and Content of Multiple-Choice Items

There are many different ways to format multiple-choice questions and answers but the most typical format is to have a question (stem), a correct answer, and incorrect (but potentially plausible) answers. One main issue when making tests is the number of choices that should be provided for each question. This can vary from as few as two (e.g., true/false questions) to as many as the test-maker decides to create (although many tests including standardized exams typically utilize 4–5 choices per question). Given that creating plausible choices consumes the time and energy of instructors, it is beneficial to know the lowest number of choices that are sufficient for testing purposes. In one study

by Baghaei and Amrahi (2011), 180 English undergraduates took three different versions of the same (30 multiple-choice items) grammar test with either five, four, or three choices per item (the four- and three-choice versions were created by randomly deleting one or two incorrect choices from the five-item version). They found no evidence of significant differences across the three versions in test reliability, response behaviors, item difficulty, or item fit statistics, and thus concluded that three items per multiple-choice question is optimal. Rodriguez (2005) conducted a meta-analysis on 27 studies (across multiple domains including subject matter and age level) on this topic and also concluded that three choices is the optimal format for multiple-choice questions. They also note that three choices may provide an additional benefit, as more three-choice questions can be administered in the same testing time-period as questions with more than three choices. Thus, instructors who utilize three-choice items can cover more content in their exams. One study for example supported this idea and found that students answered (on average) items with three choices 5 seconds faster than those with four or five choices (Schneid, Armour, Park, Yudkowsky, & Bordage, 2014).

While certain types of questions are more difficult (students routinely perform worse) than others, they do not increase discriminability between lower- and higher-performing students and thus should be avoided (Caldwell & Pate, 2013). These include questions that use negatives (e.g., "all of the following EXCEPT . . .") and "None of the Above" as an answer choice. Other studies of multiple-choice testing also recommend that "None of the Above" and composite questions such as "A and B, but not C" be omitted from answer choices (e.g., DiBattista, Sinnige-Egger, & Fortuna, 2014; Pachai, DiBattista, & Kim, 2015), or to at least emphasize (e.g., via bolding) any negative wording in the question (J. D. Hansen & Dexter, 1997). J. D. Hansen and Dexter (1997) also offered additional recommendations for wording such as avoiding multipart questions and giveaways to correct/incorrect answers (e.g., if one choice is much longer than the others, if grammar is consistent/inconsistent between the question and answer choices, if multiple incorrect answer choices have the same meaning). The authors also pointed out that "All of the Above" should be avoided as students who know that two items are correct can select this response without actually knowing that the other items are also correct. In her book on teaching, Davis (1993) also noted (in addition to the wording tips above) that choices should be parallel in structure (e.g., number of qualifiers, details) and equal in length, and that the stem of the question should be kept as short as possible (to avoid confusion) but should contain all relevant information (e.g., the student will not need to read all the choices before they can understand the question).

Order of questions can be set up in many ways including using the same order for all copies of a particular test, randomizing the order of questions (particularly simple to do for online tests), utilizing multiple versions (with different ordering) of the same test to help prevent cheating (particularly in large classes), ordering questions based on difficulty level, and ordering questions sequentially to reflect readings and lectures (e.g., early questions ask about early chapters/lectures). Ordering in terms of difficulty level (e.g., from easy to difficult, or vice versa) does not lead to differences on performance, but does lead to differences in students' perceptions of performance such that students are more optimistic if exams progress from easy to difficult (Weinstein & Roediger, 2010, 2012). In terms of ordering that affects actual (as opposed to perceived) performance, a study by Balch (1989) found that sequential ordering led to higher scores than random or chapter-contiguous (questions on the same chapter appeared together, but not sequentially, throughout the test) orders (the order of questions did not significantly affect how long it took students to complete tests). Balch suggested that this increased performance with sequential ordering is due to facilitation of student's memory based on the idea of encoding specificity (i.e., by mimicking the context and order in which content was learned, recall is enhanced). Thus, sequential ordering may be optimal for facilitating student performance.

Answer positions should be randomized when possible (e.g., computerized software can take care of this task) as test makers have a tendency to place correct answers in middle positions. This propensity is shared (or perhaps reinforced) by test takers who are more likely to select middle-positioned items (e.g., the "c" re-

sponse in a five-choice question), particularly when guessing (Attali & Bar-Hillel, 2003).

Cheating Prevention and Countermeasures

Concerns about academic dishonesty are common with the use of multiple-choice questions, as it is easier for students to copy answers (e.g., from another student, a book, or online resources) via this format than on constructed-response questions. Numerous articles in the higher education literature note this concern and pedagogy articles offering advice and methods for detecting cheating date back to at least the 1920s (e.g., Bird, 1927). Multiple articles have addressed methods to detect cheating, often through statistical analysis of the likelihood of similar answers within pairs of students (e.g., Bellezza & Bellezza, 1989; Frary, Tideman, & Watts, 1977; Harpp & Hogan, 1993; van der Linden & Sotaridona, 2004). However, these methods for detecting answer-copying are not commonly utilized within college classroom settings, presumably due to their cost in terms of time and energy (Frary, 1993). Thus, prevention of cheating may be the most effective and feasible route for test-makers. Answer copying can be reduced by using alternate test forms, with both questions and answers rearranged, as only rearranging questions does not significantly reduce copying (Houston, 1983). As students are more likely to copy from those sitting next to them (Houston, 1976), adjusting seating such as with alternation can reduce cheating. Similarly, assigned seating via seating charts may also be utilized to prevent cheating, and can also facilitate cheating detection as who sat where is recorded allowing for error-similarity analysis (Harpp & Hogan, 1993).

Academic dishonesty can be further prevented by providing an academic integrity policy that defines cheating and encourages academic integrity so that students are clear on what is expected of them (Olt, 2002; Rowe, 2004). Additionally, honor codes (at the institutional, departmental, or even course level) and academic honesty agreements that students are asked to sign (on paper or electronically) may be utilized. There is evidence that honor codes are effective at reducing the prevalence of cheating (e.g., Hutton, 2006; McCabe & Treviño, 2002; McCabe, Treviño, & Butterfield, 2001; Vandehey, Diekhoff, & LaBeff, 2007), with some evidence that longer, formal honor codes that carry consequences are more effective (Gurung, Wilhelm, & Filz, 2012). Additionally, honor systems at the institutional level can be effective at fostering academic integrity such that students from schools with traditional honor systems (compared to students from schools with nonhonor or modified systems) rate scenarios of academic dishonesty as being more dishonest and express a higher likelihood of reporting incidences of academic dishonesty (Schwartz, Tatum, & Hageman, 2013).

There are unique issues with cheating in online/distance-learning courses as instructors are not present to ensure students do not use resources (e.g., readings, notes, the Internet) to search for answers or share answers with peers (Olt, 2002; Rowe, 2004). However, even with online classes there are countermeasures that can be taken to reduce the occurrence of student cheating on multiple-choice exams. Rowe (2004) suggested drawing questions from a large question bank and randomizing question and answer order. Deferring feedback on questions until after the quiz or exam is closed will reduce the incidence of students copying or taking screenshots of answers and sharing them with other students. Instructors can also set time and access limits to exams and quizzes using many course software programs (Olt, 2002). It is also suggested that instructors change the test questions between semesters, which can prevent sharing of questions and answers between students taking the course at different times. While it may be optimal to update questions each semester, unless one is using an already established test bank, this strategy may consume significant instructor time and energy (and defeat one of the pros of using multiple-choice tests). Thus, the most feasible strategy may be to utilize some of the other suggestions outlined in this section and update test questions as frequently as the instructor's resources allow.

Finally, in an attempt to increase academic integrity, online courses may utilize "lockdown browsers," for example, Respondus LockDown Browser® (Respondus, 2015). This is ideal for controlling the virtual environment of the students while completing assignments or assess-

ments that require independent work. Several lockdown browsers can be downloaded for free by students and are user-friendly. Most are compatible with several learning platforms (e.g., Blackboard, D2L, Moodle), and provide testing environments that prevent task switching, searching screen shots, copy and paste, and so on. Some lockdown browsers offer monitoring of the student via webcam to allow for virtual proctoring (Respondus, 2015). Stack (2015) investigated the difference between students taking an online exam through the Respondus LockDown Browser and students taking a traditional in-seat exam. He noted that there were no significant differences found between the two groups (Stack, 2015). This is important considering there has been evidence for increased levels of cheating for online courses compared to traditional courses (Lanier, 2006).

Conclusion

Multiple-choice testing is a preferred assessment method for students and instructors alike, and very commonly used in college classrooms (whether in-seat or online). While there are valid concerns about the use of multiple choice, there are also ways to help mitigate against potential negatives of this method and to enhance student learning, performance, and enjoyment as well as instructor's efficiency, understanding of student outcomes, and evaluations. For example, assessment of deeper-level thinking is a common concern with multiple-choice question use. However, with appropriate formatting of multiple-choice testing it is possible to assess deeper- level thinking (Simkin & Kuechler, 2005; Tractenberg, Gushta, Mulroney, & Weissinger, 2013). Student understanding and learning can also be increased by utilizing appropriate feedback in terms of content and timing. Efficiency and quality of test administration, grading, and student learning can be increased through formatting, test construction, and cheating countermeasures. This short review is by no means an exhaustive summary of the large pedagogy literature on testing. However, we hope that our summary will encourage the utilization of empirical data in pedagogy related decision-making and be helpful to instructors as they prepare courses and assessments.

References

Attali, Y., & Bar-Hillel, M. (2003). Guess where: The position of correct answers in multiple- choice test items as a psychometric variable. *Journal of Educational Measurement, 40,* 109–128. http://dx.doi.org/10.1111/j.1745-3984.2003.tb01099.x

Baghaei, P., & Amrahi, N. (2011). The effects of the number of options on the psychometric characteristics of multiple choice items. *Psychological Test and Assessment Modeling, 53,* 192–211.

Balch, W. R. (1989). Item order affects performance on multiple-choice exams. *Teaching of Psychology, 16,* 75–77. http://dx.doi.org/10.1207/s15328023top1602_9

Begg, I., Armour, V., & Kerr, T. (1985). On believing what we remember. *Canadian Journal of Behavioural Science Review, 17,* 199–214. http://dx.doi.org/10.1037/h0080140

Bellezza, F. S., & Bellezza, S. F. (1989). Detection of cheating on multiple-choice tests by using error-similarity analysis. *Teaching of Psychology, 16,* 151–155. http://dx.doi.org/10.1207/s15328023top1603_15

Bird, C. (1927). The detection of cheating in objective examinations. *School & Society, 25,* 261–262.

Briggs, D. C., Alonzo, A. C., Schwab, C., & Wilson, M. (2006). Diagnostic assessment with ordered multiple-choice items. *Educational Assessment, 11,* 33–63. http://dx.doi.org/10.1207/s15326977ea1101_2

Butler, A. C., Karpicke, J. D., & Roediger, H. L. III. (2007). The effect of type and timing of feedback on learning from multiple-choice tests. *Journal of Experimental Psychology: Applied, 13,* 273–281. http://dx.doi.org/10.1037/1076-898X.13.4.273

Butler, A. C., & Roediger, H. L. III. (2008). Feedback enhances the positive effects and reduces the negative effects of multiple-choice testing. *Memory & Cognition, 36,* 604–616. http://dx.doi.org/10.3758/MC.36.3.604

Caldwell, D. J., & Pate, A. N. (2013). Effects of question formats on student and item performance. *Journal of Pharmaceutical Education, 77,* 71. http://dx.doi.org/10.5688/ajpe77471

Clariana, R. B., Wagner, D., & Roher Murphy, L. C. (2000). Applying a connectionist description of feedback timing. *Educational Technology Research and Development, 48,* 5–22. http://dx.doi.org/10.1007/BF02319855

Cole, R. S., & Todd, J. B. (2003). Effects of Web-based multimedia homework with immediate rich feedback on student learning in general chemistry. *Journal of Chemical Education, 80,* 1338. http://dx.doi.org/10.1021/ed080p1338

Cortright, R. N., Collins, H. L., Rodenbaugh, D. W., & DiCarlo, S. E. (2003). Student retention of course content is improved by collaborative-group

testing. *Advances in Physiology Education, 27,* 102–108. http://dx.doi.org/10.1152/advan.00041.2002

Crowe, A., Dirks, C., & Wenderoth, M. P. (2008). Biology in bloom: Implementing Bloom's Taxonomy to enhance student learning in biology. *CBE Life Sciences Education, 7,* 368–381. http://dx.doi.org/10.1187/cbe.08-05-0024

Davies, P. (2002). There's no confidence in multiple-choice testing. *Proceedings of 6th CAA Conference*. Loughborough, UK: Loughborough University. Retrieved from https://dspace.lboro.ac.uk/dspace-jspui/bitstream/2134/1875/1/davies_p1.pdf

Davis, B. G. (1993). *Tools for teaching*. San Francisco, CA: Jossey-Bass.

DiBattista, D., & Kurzawa, L. (2011). Examination of the quality of multiple-choice items on classroom tests. *Canadian Journal for the Scholarship of Teaching and Learning, 2,* 4.

DiBattista, D., Sinnige-Egger, J. A., & Fortuna, G. (2014). The "None of the above" option in multiple-choice testing: An experimental study. *Journal of Experimental Education, 82,* 168–183. http://dx.doi.org/10.1080/00220973.2013.795127

Douglas, M., Wilson, J., & Ennis, S. (2012). Multiple-choice question tests: A convenient, flexible, and effective learning tool? A case study. *Innovations in Education and Teaching International, 49,* 111–121. http://dx.doi.org/10.1080/14703297.2012.677596

Echternacht, G. T., Boldt, R. F., & Sellman, W. S. (1972). Personality influences on confidence test scores. *Journal of Educational Measurement, 9,* 235–241. http://dx.doi.org/10.1111/j.1745-3984.1972.tb00957.x

Foster, D., & Miller, H. L. (2009). A new format for multiple-choice testing: Discrete-option multiple-choice. Results from early studies. *Psychology Science Quarterly, 51,* 355–369.

Frary, R. B. (1993). Statistical detection of multiple-choice answer copying: Review and commentary. *Applied Measurement in Education, 6,* 153–165. http://dx.doi.org/10.1207/s15324818ame0602_4

Frary, R. B. (1988). Formula scoring of multiple-choice tests (correction for guessing). *Educational Measurement: Issues and Practice, 7,* 33–38. http://dx.doi.org/10.1111/j.1745-3992.1988.tb00434.x

Frary, R. B., Tideman, T. N., & Watts, T. M. (1977). Indices of cheating on multiple-choice tests. *Journal of Educational and Behavioral Statistics, 2,* 235–256. http://dx.doi.org/10.3102/10769986002004235

Funk, S. C., & Dickson, K. L. (2011). Multiple-choice and short-answer exam performance in a college classroom. *Teaching of Psychology, 38,* 273–277. http://dx.doi.org/10.1177/0098628311421329

Gaytan, J., & McEwen, B. C. (2007). Effective online instructional and assessment strategies. *American Journal of Distance Education, 21,* 117–132. http://dx.doi.org/10.1080/08923640701341653

Grühn, D., & Cheng, Y. (2014). A self-correcting approach to multiple-choice exams improves students' learning. *Teaching of Psychology, 41,* 335–339. http://dx.doi.org/10.1177/0098628314549706

Gurung, R., Wilhelm, T., & Filz, T. (2012). Optimizing honor codes for online exam administration. *Ethics & Behavior, 22,* 158–162. http://dx.doi.org/10.1080/10508422.2011.641836

Haladyna, T. M., Downing, S. M., & Rodriguez, M. C. (2002). A review of multiple-choice item-writing guidelines for classroom assessment. *Applied Measurement in Education, 15,* 309–333. http://dx.doi.org/10.1207/S15324818AME1503_5

Hansen, J. D., & Dexter, L. (1997). Quality multiple-choice test questions: Item-writing guidelines and an analysis of auditing testbanks. *Journal of Education for Business, 73,* 94–97. http://dx.doi.org/10.1080/08832329709601623

Hansen, R. (1971). The influence of variables other than knowledge on probabilistic tests. *Journal of Educational Measurement, 8,* 9–14. http://dx.doi.org/10.1111/j.1745-3984.1971.tb00900.x

Harpp, D. N., & Hogan, J. J. (1993). Crime in the classroom: Detection and prevention of cheating on multiple-choice exams. *Journal of Chemical Education, 70,* 306–323. http://dx.doi.org/10.1021/ed070p306

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research, 77,* 81–112. http://dx.doi.org/10.3102/003465430298487

Hickson, S., Reed, W. R., & Sander, N. (2012). Estimating the effect on grades of using multiple-choice versus constructive-response questions: Data from the classroom. *Educational Assessment, 17,* 200–213. http://dx.doi.org/10.1080/10627197.2012.735915

Houston, J. P. (1976). The assessment and prevention of answer copying on undergraduate multiple-choice examinations. *Research in Higher Education, 5,* 301–311. http://dx.doi.org/10.1007/BF00993429

Houston, J. P. (1983). Alternate test forms as a means of reducing multiple-choice answer copying in the classroom. *Journal of Educational Psychology, 75,* 572–575. http://dx.doi.org/10.1037/0022-0663.75.4.572

Hutton, P. A. (2006). Understanding student cheating and what educators can do about it. *College Teaching, 54,* 171–176. http://dx.doi.org/10.3200/CTCH.54.1.171-176

Jacobs, S. S. (1971). Correlates of unwarranted confidence in responses to objective test items. *Jour-*

nal of Educational Measurement, 8, 15–20. http://dx.doi.org/10.1111/j.1745-3984.1971.tb00901.x

Jensen, M., Duranczyk, I., Staat, S., Moore, R., Hatch, J., & Somdahl, C. (2006). Using a reciprocal teaching strategy to create multiple-choice exam questions. *American Biology Teacher, 68,* 67–71. http://dx.doi.org/10.1662/0002-7685(2006)68[67:TCMEQ]2.0.CO;2

Kapitanoff, S. H. (2009). Collaborative testing: Cognitive and interpersonal processes related to enhanced test performance. *Active Learning in Higher Education, 10,* 56–70. http://dx.doi.org/10.1177/1469787408100195

Kulik, J. A., & Kulik, C. C. (1988). Timing of feedback and verbal learning. *Review of Educational Research, 58,* 79–97. http://dx.doi.org/10.3102/00346543058001079

Landrum, E. R., & Gurung, A. R. (2013). The memorability of introductory psychology revisited. *Teaching of Psychology, 40,* 222–227. http://dx.doi.org/10.1177/0098628313487417

Lanier, M. M. (2006). Academic integrity and distance learning. *Journal of Criminal Justice Education, 17,* 244–261. http://dx.doi.org/10.1080/10511250600866166

Marsh, E. J., Roediger, H. L. III, Bjork, R. A., & Bjork, E. L. (2007). The memorial consequences of multiple-choice testing. *Psychonomic Bulletin & Review, 14,* 194–199. http://dx.doi.org/10.3758/BF03194051

McCabe, D. L., & Treviño, L. K. (2002). Honesty and honor codes. *Academe, 88,* 37–41. http://dx.doi.org/10.2307/40252118

McCabe, D. L., Treviño, L. K., & Butterfield, K. D. (2001). Cheating in academic institutions: A decade of research. *Ethics & Behavior, 11,* 219–232. http://dx.doi.org/10.1207/S15327019EB1103_2

McCoubrie, P. (2004). Improving the fairness of multiple-choice questions: A literature review. *Medical Teacher, 26,* 709–712. http://dx.doi.org/10.1080/01421590400013495

Montepare, J. M. (2005, October). A self-correcting approach to multiple choice tests. *APS Observer, 18,* 35–36.

Montepare, J. M. (2007). A self-correcting approach to multiple choice tests. In B. Perlman, L. I. McCann, & S. H. McFadden (Eds.), *Lessons learned* (Vol. 3, pp. 143–154). Washington, DC: Association for Psychological Science.

Olt, M. R. (2002). Ethics and distance education: Strategies for minimizing academic dishonesty in online assessment. *Online Journal of Distance Learning Administration, 5*(3).

Ozuru, Y., Briner, S., Kurby, C. A., & McNamara, D. S. (2013). Comparing comprehension measured by multiple-choice and open-ended questions. *Canadian Journal of Experimental Psychology, 67,* 215–227. http://dx.doi.org/10.1037/a0032918

Pachai, M. V., DiBattista, D., & Kim, J. A. (2015). A systematic assessment of 'none of the above' on multiple choice tests in a first year psychology classroom. *Canadian Journal for the Scholarship of Teaching and Learning*. Advance online publication. http://dx.doi.org/10.5206/cjsotl-rcacea.2015.3.2

Park, J. (2010). Constructive multiple-choice testing system. *British Journal of Educational Technology, 41,* 1054–1064. http://dx.doi.org/10.1111/j.1467-8535.2010.01058.x

Pridemore, D., & Klein, J. D. (1991). Control feedback in computer-assisted instruction. *Educational Technology Research and Development, 39,* 27–32. http://dx.doi.org/10.1007/BF02296569

Rao, S. P., Collins, H. L., & DiCarlo, S. E. (2002). Collaborative testing enhances student learning. *Advances in Physiology Education, 26,* 37–41. http://dx.doi.org/10.1152/advan.00032.2001

Respondus. (2015, November 2). *LockDown Browser instructor resources*. Retrieved from https://www.respondus.com/products/lockdown-browser/resources.shtml

Rodriguez, M. C. (2005). Three options are optimal for multiple-choice items: A meta-analysis of 80 years of research. *Educational Measurement: Issues and Practice, 24,* 3–13. http://dx.doi.org/10.1111/j.1745-3992.2005.00006.x

Roediger, H. L. III, & Marsh, E. J. (2005). The positive and negative consequences of multiple-choice testing. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31,* 1155–1159. http://dx.doi.org/10.1037/0278-7393.31.5.1155

Rowe, N. C. (2004). Cheating in online student assessment: Beyond plagiarism. *Online Journal of Distance Learning Administration, 7*(2).

Schneid, S. D., Armour, C., Park, Y. S., Yudkowsky, R., & Bordage, G. (2014). Reducing the number of options on multiple-choice questions: Response time, psychometrics and standard setting. *Medical Education, 48,* 1020–1027. http://dx.doi.org/10.1111/medu.12525

Schwartz, B., Tatum, H., & Hageman, M. (2013). College students' perceptions of and responses to cheating at traditional, modified, and non-honor system institutions. *Ethics & Behavior, 23,* 463–476. http://dx.doi.org/10.1080/10508422.2013.814538

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

Scouller, K. M., & Prosser, M. (1994). Students' experiences in studying for multiple choice question examinations. *Studies in Higher Education,*

19, 267–279. http://dx.doi.org/10.1080/03075079412331381870

Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research, 78,* 153–189. http://dx.doi.org/10.3102/0034654307313795

Simkin, M., & Kuechler, W. L. (2005). Multiple-choice tests and student understanding: What is the connection? *Decision Sciences Journal of Innovative Education, 3,* 73–98. http://dx.doi.org/10.1111/j.1540-4609.2005.00053.x

Stack, S. (2015). The impact of exam environments on student test scores in online courses. *Journal of Criminal Justice Education, 26,* 273–282. http://dx.doi.org/10.1080/10511253.2015.1012173

Steven Pinker's mind games. (2014, April 11). *New York Times*. Retrieved January 9, 2016, from http://www.nytimes.com/interactive/2014/04/13/education/edlife/edlife-quiz-psych.html?_r=0

Struyven, K., Dochy, F., & Janssens, S. (2005). Students' perceptions about evaluation and assessment in higher education: A review. *Assessment & Evaluation in Higher Education, 30,* 325–347. http://dx.doi.org/10.1080/02602930500099102

Tozoglu, D., Tozoglu, M. D., Gurses, A., & Dogar, C. (2004). The students' perceptions: Essay versus multiple-choice type exams. *Journal of Baltic Science Education, 2,* 52–59.

Tractenberg, R. E., Gushta, M. M., Mulroney, S. E., & Weissinger, P. A. (2013). Multiple choice questions can be designed or revised to challenge learners' critical thinking. *Advances in Health Sciences Education: Theory and Practice, 18,* 945–961. http://dx.doi.org/10.1007/s10459-012-9434-4

Vandehey, M. A., Diekhoff, G. M., & LaBeff, E. E. (2007). College cheating: A twenty-year follow-up and the addition of an honor code. *Journal of College Student Development, 48,* 468–480. http://dx.doi.org/10.1353/csd.2007.0043

van der Linden, W. J., & Sotaridona, L. (2004). A statistical test for detecting answer copying on multiple-choice tests. *Journal of Educational Measurement, 41,* 361–377. http://dx.doi.org/10.1111/j.1745-3984.2004.tb01171.x

Weinstein, Y., & Roediger, H. L. III. (2010). Retrospective bias in test performance: Providing easy items at the beginning of a test makes students believe they did better on it. *Memory & Cognition, 38,* 366–376. http://dx.doi.org/10.3758/MC.38.3.366

Weinstein, Y., & Roediger, H. L. III. (2012). The effect of question order on evaluations of test performance: How does the bias evolve? *Memory & Cognition, 40,* 727–735. http://dx.doi.org/10.3758/s13421-012-0187-3

Yonker, J. E. (2011). The relationship of deep and surface study approaches on factual and applied test-bank multiple-choice question performance. *Assessment & Evaluation in Higher Education, 36,* 673–686. http://dx.doi.org/10.1080/02602938.2010.481041

Zeidner, M. (1987). Essay versus multiple-choice type classroom exams: The students' perspective. *Journal of Educational Research, 80,* 352–358. http://dx.doi.org/10.1080/00220671.1987.10885782

Zimmaro, D. W. (2010). *Writing good multiple-choice exams* [Workshop material]. Austin, TX: Learning Sciences, University of Texas—Austin. Retrieved from https://learningsciences.utexas.edu/sites/default/files/writing-good-multiple-choice-exams-04-28-10_0.pdf

Received November 4, 2015
Revision received May 5, 2016
Accepted May 9, 2016 ■