

# **Constructive Response Vs. Multiple-Choice Tests In Math: American Experience And Discussion (Review)**

*Nina V. Stankous, PhD*  
National University, CA, USA

---

## **Abstract**

This study investigates and compares multiple-choice and constructive-response test formats for assessments in mathematics. This topic has been researched and debated for many years by students, educators, and politicians. Even between students there is a wide range of disagreement between which test format is preferred. Prior research has revealed both benefits and disadvantages for both formats. The study found that although multiple-choice tests are often used to measure student and teacher performance in math education, constructive-response test formats better represent student learning. Some factors that will be discussed are the “guessing” problem, partial credit, grading efficiency, and the ability to cover topics.

---

**Keywords:** Test formats, mathematics

## **Introduction**

The question over how we measure student performance is one that has been debated consistently since the conception of formal testing. As a nation, the United States of America has sought to determine a means to measure student and teacher performance in a way that is fair to teachers yet an accurate reflection of student learning. Although there are many different methods of testing, this review will focus on multiple-choice and constructive-response test formats. These two test formats are generally the most popular and sit at opposite ends of the spectrum; therefore, they have generated the most interest and research.

Multiple-choice tests are often selected due to their ease of grading, grading consistency, ability to cover a vast amount of topics, and for their generally high-performing results. Teachers who are evaluated or incentivized by student test scores might prefer to administer an “easier” multiple-choice test that tends to give higher scores than a constructive-response format.

Constructive-response tests are unique in their ability to award partial credit for incorrect answers and are generally considered the “harder” tests due to the inability to guess. Researchers have tried to determine the relationship between learning and test scores, but often it is hard to determine. Because multiple-choice tests usually produce higher test scores, student achievement is often considered equal or higher than constructive-response versions. This debate has carried over into standardized testing, testing throughout the term, and college entrance exams.

### **Purpose of the review**

I hope this review will help other educators think about using effective test formats to accurately measure student performance. The main questions are the following:

1. What are the factors that influence a teacher’s decision to use each test format?
2. How do multiple-choice and constructive-response test formats affect student learning and retention?
3. Is a multiple-choice or constructive-response test format better to measure student learning and teacher performance?
4. What are solutions for teachers to be able to administer the most effective test formats?

By investigating and gathering the research related to this topic, we can make conclusions about how these assessment formats affect student learning. It is important to investigate the procedures used for each study as well as the results in order to make informed conclusions about the topics. If we investigate the factors that influence an educator’s decision to use each test format that will provide insight into their research. It is also important that once the best test format has been determined, solutions are discussed that make administering that test an achievable

### **Resources**

There have been many studies conducted measuring different effects multiple-choice and constructive-response test formats have on students. Many of these studies have followed the same general set-up: give a set of students a series of either multiple-choice or constructive-response tests and compare the results. The true difficulty lies in the fact that it is not an exact science. There is research to support both sides of the argument and this is why there is so much debate in the education system.

One problem with some of these studies is that the subjects were not math students. I wonder if those research studies would have more dramatic results if they were measuring performance on math tests because of the ability to work backwards. In many other subjects, if a student were blindly

picking multiple-choice items, (without any knowledge of the subject) he would probably receive a better score than a student in the same situation taking a free-response test. However in math it is often possible to work backwards and arrive at the correct answer without knowing how to do the problem. In this situation the scores might be significantly different, with a possibility for a perfect score without knowing the material for the multiple-choice version. The issue of whether or not multiple-choice and constructive-response test formats are better has been debated and studied, but this issue specifically in a math education context has not been researched thoroughly enough.

In the Chaoui (2011) study, participants were given the mock California High School Exit Exam (CAHSEE) in mathematics in both a multiple-choice format and a constructive-response format. The results were then compared using t-tests and other statistical analysis. The participants were freshman and sophomore algebra I, algebra II, and geometry students. The questions were selected so that they reflected standards from Number Sense, Statistics and Probability, Algebra and Functions, Algebra One, and Measurement and Geometry.

The Gay (1980) study was interesting because, although the subjects were not math students, the study measured the ability to recall information on both multiple-choice and constructive-response tests, when the students had received either multiple-choice or constructive-response tests up until that point. The mean test scores and standard deviations were found and compared with each other. The participants in this study were educational research students at Florida International University. Test questions measured equivalent concepts involving knowledge, comprehension, and application items.

The Elbrink and Waits (1970) study participants were Calculus II students at Ohio State University. The entire population of students enrolled in the course was split into two groups where one half was given multiple-choice examinations and the other half was given constructive-response examinations. The questions on the examinations were identical except for the method of response. An analysis of covariance was used to find differences in the effects of mathematical achievement.

The Frary (1985) simulation study was not performed on actual students, but simulated using a computer program. “The simulation was replicated three times for each of 30 variations reflecting format and the extent to which examinees were (a) misinformed, (b) successful in guessing constructive-response answers, and (c) able to recognize with assurance correct multiple-choice options that they could not produce under constructive-response testing” (p.21).

The participants of the O’Neil and Brown (1997) study were eighth grade math students. The length of the study was relatively short, with only seven multiple-choice items and one constructive-response question. Although the length of the study was short, the primary objective of the study was to research the metacognitive effects of math assessments and not as focused on measuring mathematical achievement. The study was still able to accomplish its goal through the attached questionnaires that each student answered.

## **Discussion**

Multiple-choice tests are often selected because of the ease of grading. Teachers can have students use scantrons and then use a machine (provided by the school) to grade for automated grading. Each question is either right or wrong, there is no partial credit. This significantly cuts down the time each teacher needs to spend grading exams. Another reason why multiple-choice tests have been so popular is for grading consistency. Because there is no partial credit and the correct answer is either selected or not selected, it makes grading tests a very impartial task. Other formats leave room for interpretation and for inconsistent grading (Chaoui, 2011).

Teachers who are evaluated or incentivized by their students’ test scores might prefer to administer an “easier” multiple-choice test that tends to give higher scores than a constructive-response format. This leads to a culture of apathy towards meaningful testing. Chaoui (2011) states, “since teachers are being held accountable for their teaching by virtue of their test scores, they may prefer to give the students tests on which they are more likely to be successful” (p.128).

The ability for a student to guess can be both an advantage and disadvantage for multiple-choice tests. On one hand, the student will have a higher score on a problem they have no idea how to solve. On the other hand, it is not very reflective of the student’s knowledge on the subject and provides a false sense of retention. Yet, the scores are usually higher than the constructive-response questions, so many teachers prefer it. Since teachers are evaluated based on the performance of their students, it is understandable why they would prefer to give a test they believe they will score higher on. Chaoui suggests that “Integrating open-ended math problems, as well as implementing performance tasks, which promote cognitive thinking, will prepare the students to be more confident and efficient problem solvers. Teachers must strive to incorporate multiple-choice and constructed response items on their tests to assess skills as well as literacy” (Chaoui, 2011, p.130).

Constructive-response formatted test questions can vary in difficulty, much like multiple-choice test questions. A criticism of constructive-response tests is that you cannot ask as many questions, because they take

longer for the students to complete. Chaoui (2011) states “Part of effective instruction is giving students opportunities to explain their thinking in writing, using proofs, multiple steps, organizers and written sentences” (p.9). Many instructors in math education focus on the students’ problem solving abilities and do not focus enough on explanation. If it is our goal, as educators, for students to retain the information they are taught, then our primary focus should not be on the students’ ability to score well on a test. Rather, our focus should be on their ability to consistently perform these same tasks in the future and be able to explain how to perform these operations to others.

Sometimes wrong answers tell us more about a student’s learning than the right answers. When a student selects an incorrect answer on a multiple-choice test it doesn’t tell you very much about how they got to that answer. In a constructive-response test, however, you can often see exactly where the student went wrong. “Open-ended questions provide insights into the misconceptions of students and allow the teacher to evaluate the various techniques they use” (Chaoui, 2011, p.18). Not only are seeing mistakes beneficial to an instructor, but it is very advantageous to the student as well. If the student is able to see exactly what they were thinking while they were testing, they have a better chance of learning from their mistake. When they try to recall that information later on, they will think about the mistake they made the last time and be less likely to make the same mistake again.

A disadvantage to constructive-response tests is the extra time it takes to grade. Because each response may be slightly different it takes more time to assign partial credit, which also leads to grading inconsistencies. Various follow up procedures would need to be in place to ensure fairness, which would be expensive.

Another issue with multiple-choice tests is the distracter element. When instructors are trying to make their tests harder they will sometimes add an incorrect answer that may be close to the correct answer to the list of available choices. This may cause a student to select a wrong answer when they might have solved the problem correctly without that distracter, in a constructive-response format (Frery, 1985, p.26). This also may lead to incorrect retention. One study found that “Multiple-choice testing may inadvertently lead to the creation of false knowledge” (Roediger, 2005, p.1156).

Another question that is raised when comparing these two test formats is the type of thinking that is involved in solving the problems. “Multiple-choice (MC) tests are depicted as assessing simple factual recognition, and constructive-response (CR) tests are depicted as evaluating higher order thinking”(Chaoui, 2011, p.1). This leads to the question about the purpose of education and educational psychology: Is the purpose of

education to get the student to advance along their educational path, or is the purpose to instill knowledge that they will be able to recall years later? The student should be able to answer questions not because they were able to memorize the answer, but because they understood what it really meant or why they learned it.

Many times assessments are compared based on their reliability and validity. In theory, a good test is defined as both reliable and valid. In actuality, a good test is difficult to develop. “Test validity refers to the degree to which the test actually measures what it claims to measure. It is the extent to which inferences, conclusions, and decisions, made on the basis of test scores are appropriate and meaningful” (Chaoui, 2011, p. 8). Chaoui argues that validity is not harmed by educated guesses because it involves narrowing choices down based on some knowledge of the material. Test reliability concerns a test’s ability to measure what is intended to be measured, or the purpose of the test.

The Frary (1985) simulation study found that both reliability and validity of constructive-response scores were higher than multiple-choice scores, and therefore recommended that constructive-response tests be used over multiple-choice tests in certain settings. It was also concluded in this study that “the potential gain might be considered too small to warrant the use of constructive-response tests, given factors such as labor required to score them” (p.31). This suggested that the extra time required to grade constructive-response tests might not be worth it.

In the Elbrink and Waits (1970) study at Ohio State University, students taking multiple-choice math tests outperformed students taking constructive-response math tests in the same subject. It was concluded that although multiple-choice tests were easier than constructive-response tests, they were just as effective in “evaluating students’ mathematical achievement” (p.4). The authors made this conclusion after taking into account the ability to guess, however their definition of “mathematical achievement” is essentially the student’s ability to problem solve, i.e. not their ability to explain thinking or prove concepts.

In the Gay (1980) study, the researcher tried to determine which testing method resulted in better retention. The study found that students who had been used to taking constructive-response tests performed just as well on the multiple-choice items as the students that had been used to taking the multiple-choice tests. Yet, the constructive-response students tested much better on the constructive-response items than the multiple-choice students. This implies that constructive-response testing results in better retention because otherwise, the multiple-choice students would have performed just as well or better on the constructive-response items.

From the O’Neil and Brown (1997) study, differential effects of multiple-choice and constructive-response test formats in math assessment on metacognition and effect was approached. The study found that constructive-response questions induced more cognitive strategy usage than multiple-choice questions. Meaning students thought about their own thinking to develop strategies to solve the problems more in constructive-response test questions than in multiple-choice test questions.

## **Results**

Although the magnitude of the results was often small, the results support my hypothesis that constructive-response tests promote student learning better than multiple-choice tests.

In the Gay (1980) study, where retention was measured, she found that “the results of this study indicate that [short-answer] testing results in equal or greater retention than [multiple-choice] testing, depending upon the mode of retention testing” (p.50). Students who had been used to taking constructive-response tests performed just as well on the multiple-choice items as the students that had been used to taking the multiple-choice tests, yet the constructive-response students tested much better on the constructive-response items than the multiple-choice students.

Educators should want meaningful test scores from their students in order to be better teachers and help students learn. constructive-response test scores are simply more meaningful than multiple-choice test scores. The Frary (1985) simulation study found that both reliability and validity of constructive-response scores were higher than multiple-choice scores, and therefore recommended that constructive-response tests be used over multiple-choice tests in certain settings.

According to the National Council of Teachers of Mathematics (1991), “although the commonly used [multiple-choice] format may yield important data, it can have a negative impact on how students are taught and evaluated at the school level because: a) Student scores are generated solely on the basis of right and wrong answers with no consideration or credit given to students’ strategies, b) Routine timing measures how quickly students can respond but not necessarily how well they think, and c) Mathematics tools such as calculators and measurement devices are not permitted” (p.8).

Although there has not been extensive research in this field, the research we conduct suggests there is a relationship between greater learning and constructive-response test formats. The strength of that relationship is debatable and depends on the factors you consider.

We need to define success in more ways than the ability to problem solve in mathematics education. We cannot measure that success from multiple-choice tests alone. Although making more constructive-response

tests would add a significant amount of time to the grading process, it would give us a more accurate depiction of how our students are performing. With the new technologies in grading free-response math answers, soon it may be possible to grade them even faster.

An important factor of this topic is in the difference between long term and short-term learning. Although a student may be able to select the correct answer on a multiple-choice question, this may be due to short-term memorization rather than long-term understanding of the topic. In fact, many schools teach to the test specifically for this reason. They want students to recognize questions or question types and memorize the correct answers so that they can meet certain educational performance standards. But once these tests are over, the student does not retain the information for a reasonable amount of time.

As an educator, your overall goal is that a student leaves your class having learned something. Many teachers believe that a student can greatly benefit from seeing exactly where they went wrong on a question. This way when you try to recall that information later, you remember that mistake and you are careful not to make it again. This is a huge advantage to constructive-response tests.

### **References:**

- Birenbaum, M., & Tatsuoka, K.K. (1987). Open-ended versus multiple-choice response formats-It does make a difference for diagnostic purposes. *Applied Psychological Measurement*, 11, 385-395.
- Bridgeman, B. (1992). A comparison of quantitative questions in open-ended and multiple-choice formats. *Journal of Educational Measurement*, 29(3), 253-271.
- Chaoui, N (2011) "Finding Relationships Between Multiple-Choice Math Tests and Their Stem-Equivalent Constructed Responses" (2011). *CGU Theses & Dissertations*. Paper 21.
- Elbrink, L., & Waits, B. (Spring, 1970). A Statistical Analysis of Multiple-Choice Examinations in Mathematics. *The Two-Year College Mathematics Journal*, 1(1), 25-29.
- Frary, R. (Spring, 1985). Multiple-Choice Versus Free-Response: A Simulation Study. *Journal of Educational Measurement*, 22, 21-31.
- Gay, L. (Spring, 1980). The Comparative Effects Of Multiple-Choice Versus Short-Answer Tests On Retention. *Journal of Educational Measurement*, 17(1), 45-50.
- National Council of Teachers of Mathematics (1991). *Mathematics Assessment: Myths, Models, Good Questions, and Practical Suggestions*. Reston, VA.

O'Neil Jr., H., & Brown, R. (1997). Differential Effects Of Question Formats In Math Assessment On Metacognition And Affect. *Applied Measurement in Education*, 331-351.

Roediger III, Henry L., Marsh, Elizabeth J. (September, 2005), The Positive and Negative Consequences of Multiple-Choice Testing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, Vol 31(5), 1155-1159.

Thissen, D., Wainer, H., & Wang, X. (Summer, 1994). Are Tests Comprising Both Multiple-Choice and Free-Response Items Necessarily Less Unidimensional Than Multiple-Choice Tests? An Analysis of Two Tests. *Journal of Educational Measurement*, 31(2), 113-123.