COMPUTER BASED TESTING COMPARED TO PAPER AND PENCIL TESTING IN

THIRD GRADE MATH

By

CAROLYN GRAY

Submitted to

The Educational Leadership Faculty
Northwest Missouri State University
Department of Professional Education
College of Education and Human Services
Maryville, MO 64468

Submitted in Fulfillment for the Requirements for

61-683 Research Paper

Summer 2012

Submitted July 13, 2013
ABSTRACT

Computer-based tests are becoming more and more common, especially due to the fact that the new Common Core State Standards will be tested using Computer Adaptive Testing methods. This study compared one third grade class’s scores on a paper-based test with an equivalent computer-based tests. Students were given two weeks between versions of test and were not given any feedback on how they did on one version before they took the second version. No statistically significant difference was found between the two modalities of tests in this study. However, students struggled with the computerized tests and needed additional instruction in how to take that mode of test. Further studies are needed to know exactly what type and amount of instruction is optimal before students are asked to take a computerized test.
Introduction

Background, Issues and Concerns

According to the No Child Left Behind Act (NCLB) of 2001 (ED.gov, n.d.), all schools that receive federal funding must administer a standardized test to ensure student achievement and improvement over time. These tests must be based on the standards teachers are required to teach and should show high achievement for all sub-populations, including low socio-economic status and varying ethnicities.

Many criticisms of the NCLB Act argue that each state’s standards are different, and we cannot compare scores across state lines. They also assert that students moving from one state to another might experience gaps or redundancies in their education due to differing curriculums. Critics claim that the NCLB Act forces schools to teach primarily “at risk” students for fear of failing the standardized tests, but not to actually prepare all students to be ready for life in an internationally competitive society. Because of these concerns, many states are switching from state standards and tests to a national set of standards, called the Common Core State Standards (CCSS). The goal of these standards is to focus on college and career readiness.

Missouri is one of forty-five states that have adopted the new CCSS. With the new standards, come new standardized tests. There are two testing consortiums creating tests to assess the Common Core Standards. Twenty states, including Missouri, have joined the Smarter Balanced Consortium as governing states, and an additional four are advisory states (Smarter Balance Assessment Consortium, 2012).

Missouri Schools will replace the Missouri Assessment Program (MAP) test with the Smarter Balance Assessment as the standardized test for public schools in the 2014-2015 school year. The Smarter Balance Assessment will be a computer-based test, where the MAP test was a
paper and pencil test. Some parents, teachers, and administrators are concerned that a new testing format will negatively affect students’ scores.

Practice under Investigation

Online testing methods were investigated to determine if students perform differently when assessed on a computer instead of on paper.

School Policy to be Informed by Study

After this study, schools will know if there are benefits or detriments to converting to computer based testing. If there are negative aspects to computerized testing, further studies can be conducted to determine what steps instructors can take to negate those aspects.

Conceptual Underpinning

The Smarter Balance Assessment Consortium has chosen to develop computer-based tests to implement Computer Adaptive Testing (CAT). The Consortium feels that CAT assessments provide faster and more accurate results, and also prevent cheating, since students are unlikely to be answering the same question simultaneously. The Smarter Balance Assessment will also contain performance tasks and extended response items. This correlates to the Framework for 21st Century Learning from the Partnership for 21st Century Skills (www.p21.org), which emphasizes use of media in critical thinking and communication tasks. CAT is not expected to affect students’ achievement levels.

Statement of the Problem

Students often struggle with assignments on the computer due to lack of computer knowledge, unrelated to subject-matter knowledge. If assessments are required to be computerized, some students might see unnecessarily low test scores due to their lack of computer literacy.
Purpose of the Study

The purpose of this study is to determine if there is a difference between student achievement using assessments on the computer compared to traditional paper and pencil tests.

Research Question

RQ: Is there a difference in student achievement between assessments on the computer and traditional paper and pencil testing?

Null Hypothesis

Ho: There is no difference in student achievement between assessments on the computer and traditional paper pencil testing.

Anticipated Benefits of the Study

This study should show the difference in student achievement when using computer based tests compared with paper and pencil assessments.

Definition of Terms

Common Core State Standards (CCSS): These are national standards for what students should be able to know and do by the end of a grade level or course. Many states, including Missouri, are adopting these standards in lieu of their own state standards.

Smarter Balance: A national standardized test aligned to the CCSS that the state of Missouri will adopt for the 2014-2015 school year. There will be tests for English language arts and mathematics for grades 3-8 and 11. This test will use computer adaptive testing and include extended response questions and performance task items, rather than only multiple-choice questions.
Missouri Assessment Program (MAP test): The current standardized test for the state of Missouri. There are communication arts tests and mathematics tests for students in grades 3-8, as well as science tests for grades 5, 8, and 11.

CFA (Continuing Formative Assessment): A series of the same or similar tests, starting with a pretest and ending with a posttest, designed to measure students’ progress throughout a unit of instruction.

QuizStar: (www.quizstar.4teachers.org) An online program that allows teachers to input questions and students to sign in and answer those questions. The system allows teachers to analyze data from students’ scores and to print individual student pages.

Computer Adaptive Testing (CAT): A type of computerized test where an incorrect response leads the next question to be an easier level, while a correct response initiates a harder next question.

Summary

Some parents, teachers, and administrators are concerned that a new computerized format for standardized tests will negatively affect students’ scores. This study will determine if there is a difference between student achievement on computer based tests and paper and pencil tests. Based on the results of this study, teachers and administrators will know if further studies should be conducted to determine the best way to prepare students to transition to computerized testing methods.
Review of Literature

Today’s classroom environments have been changing in an attempt to keep the digital generation of students engaged in productive learning. The Partnership for 21st Century Learning was formed to help teachers prepare their students for twenty-first century tasks by promoting critical thinking and problem solving, collaboration, communication, and creativity and innovation, in addition to the traditional core subjects (Magner, T., Soulé, H., & Wesolwski, K., 2011). Dreon, Kerper, and Landis (2001) attest that if we want to reach today’s students, we must broadcast on their frequency, but integrating technology into school projects. They studied Tyler Binkley and how he was able to effectively teach math to his students using digital storytelling and YouTube.

Liat Eyal (2012) calls a teacher’s understanding of technology-based assessments *digital assessment literacy*. Just as we need to allow students to use technology in their learning, teachers can and should use technology to be more efficient. Computerized testing and databases can help organize data so teachers can effectively adapt their teaching strategies to fit their students’ instructional needs.

One important way that teachers can improve their digital literacy is for teacher education programs to increase the amount of instruction they provide to pre-service teachers. One reason Tyler Binkley was able to create videos and publish them to YouTube was the education he received prior to teaching (Dreon, Kerper, & Landis, 2001). Also, more and more districts are trying to incorporate technology into their professional development for currently practicing teachers. Web 2.0 tools such as blogs or wikis, interactive whiteboards, and increased use of online research are frequently topics of instructing teachers.
There is still some ongoing debate about whether or not assessments should be completely shifted online. The arguments against computerized tests include schools’ limited access to technology and cost of acquiring all of the technology necessary. Since the concept of computer-based tests in elementary is relatively new, there is little research about the best methods of instructing and administering for these new tests (Hovland, 2005). Computers can offer opportunities for accessibility for students with disabilities, but there are also added challenges for developing accessible computerized tests. Thompson, Thurlow, and Moore (2003) offer several suggestions for accommodations for computer-based assessments, in presentations, timing, setting, and response options.

Those arguing for computer based testing state that with the Computer Adaptive Testing (CAT) methods, students will be less able to cheat and their scores will be more accurate. Also, because computers can grade the tests as they are taken, students and teachers receive results much faster than with paper versions, and with less cost of shipping materials. Hovland (2005) believes that the positives of computerized testing outweigh the negatives, and that paper-based tests will soon become “obsolete.”

A few studies have been conducted to compare computer-based tests to paper-based ones. Michael S. Page (2002) studied ten classes of third and fifth graders from low socio-economic backgrounds, half of which were given a technology-enriched curriculum. He found no significant difference between pretest and posttest scores in reading, but a significant increase in math scores for the experimental group. Also, the technology-enriched classes had indications of higher student self-esteem, as evidenced by more student-initiated conversations. Mark Pomplun has done several studies comparing computerized and paper tests. In his 2006 study, Pomplun determined that students from a lower socio-economic status were more likely to have lower
scores on computerized tests, but this disparity reduced as students moved into higher grades, most likely due to increased familiarity with computers in the school setting. In his follow-up 2007 study, Pomplun determined that both versions of tests offered equivalent assessments of reading skills, but he recommended further studies into other subject areas.

Computer-based tests are being used more and more frequently, so teachers and students will have to adapt to them whether they want to or not. Important things to remember are to educate teachers in instructing with technology, have instructors explicitly teach strategies in computer test taking skills, and allow students time for practicing with technology. In addition, it is important to remember to use technology, not just for the sake of saying you are doing so, but to increase student engagement and learning (Wilson, Wright, Inman, & Matherson, 2011).
Research Methods

Research Design

The researcher’s class contained twenty-six third grade students. Starting in October, the students took a similar version of a math CFA each month. It contained two questions on numbers and equivalent representations, three questions on missing addends, two estimation questions, two graphing questions, two elapsed time questions, two geometry questions, and one complex word problem. Students were scored on a scale of one to four for each standard. For most standards, the highest possible score was a three, but students could have earned higher than a three by showing above grade-level understanding on the missing addends, elapsed time, and complex word problem.

The instructor input the tests into QuizStar exactly the way they looked on paper. For the February test, students were asked to take the computerized test first, then the paper-based test two weeks later. On the March CFA, students took the paper test first, then the computer-based test two weeks later. For both administrations, the teacher would not help with any specific questions from students, only with reading one word per question (if requested). For all tests, students were monitored to make sure they were not communicating with others, and they were encouraged to write out their work and show how they reached their answers. Also, students were not given any feedback on their scores on one version before taking the second version of the test.

The students’ average scores on the paper-based and computer-based tests were compared using a t-test to see if there was any preference for one type of assessment. The
independent variable was whether the test taken was on paper or on the computer, and the
dependent variable was the student’s score.

Study Group Description

Twenty-six third grade students were studied from a suburban school district. There were
thirteen girls and thirteen boys. Six students qualified for free or reduced lunch. Three students
were enrolled in the gifted and talented program, and five students were considered “at risk”
academically. Students came from a variety of socioeconomic status backgrounds and had
varying levels of experience on computers.

Data Collection and Instrumentation

The CFAs used for this experiment were teacher created and have been used for the past
8 years. The assessments were inputted to the computer exactly the way they were on the paper
and pencil tests. Students were encouraged to use scratch paper to show work for both versions
of the CFAs. The students took a similar version of a test each month during the school year,
starting in October. The March and April tests were taken both on the computer and on paper,
and the differences in scores were noted.

Statistical Analysis Methods

To analyze the collected data, students average scores were calculated in Microsoft
Excel. Then, a t-test was used to compare the paper-based test and computer-based test averages.
The t-test was completed using A Statistical Package (ASP) software.
Findings

Due to computer difficulty, a lack of students understanding how to complete the test on the computer, and many students not being able to finish the test, the February CFA scores were not regarded in the results. More information will be listed in the Conclusions section of this paper.

Before the March CFA, the instructor made sure to model answering questions on the computer and using scratch paper. Students practiced this skill in a whole-class setting before being asked to complete the second computer-based test.

A t-test was performed on the March CFA scores to determine if there was a significant difference in student scores on the computer-based test compared with the paper-based test.

<table>
<thead>
<tr>
<th>Source</th>
<th>Mean</th>
<th>Mean D</th>
<th>t-test</th>
<th>df</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Paper-based test (n=26)</td>
<td>2.75</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Computer-based test (n=20)</td>
<td>2.73</td>
<td>0.02</td>
<td>0.14</td>
<td>44</td>
<td>0.89</td>
</tr>
</tbody>
</table>

Note: Significant when p<=0.25

In the data studied, 26 students took the paper-based test. Only 20 students completed the computer-based test. The mean of the paper version was 2.75. The mean of the computer version was 2.73. The difference between the means (Mean D) was 0.02, and the t-test calculation was 0.14. The degrees of freedom (df) was 44, and the p-value was 0.89. The alpha level was set at 0.25.
The null hypothesis of this t-test comparison is that there is no significant difference between paper-based and computer-based test cores. Because the p-value of 0.89 is not lower than the alpha level of 0.25, there is no significant difference between scores when students take different versions of a test. The null hypothesis is not rejected in this study.
Conclusions and Recommendations

Because the null hypothesis was not rejected, there is no significant difference in student performance on computer-based tests compared to paper versions. That is important to note for teachers concerned with differences in scores when converting to the Computer Adaptive Testing of the Smarter Balance assessment.

However, the fact that the February computerized scores had to be thrown out is also important. When the third graders first used the computer to take their test, there was a significant amount of confusion and frustration in the classroom. Students were not using the strategies they had been taught of showing work and trying to do all of the calculations in their heads, even though they had scratch paper to work on. They were also getting very frustrated when they wanted to go back to a previous question or skip the one they were on to come back to later. It took students much longer to complete the computerized test than the paper one, because they had to figure out the control buttons on the computer.

The Smarter Balance Assessment Consortium has chosen to develop computerized testing rather than paper based tests due to increased security, decreased cost, and no difference in student achievement. This study has shown that when students are prepared to take a test online, the mode of testing does not affect achievement. However, based on February testing experience, students can experience added frustrations on a computerized test by being unfamiliar with the controls.

The recommendation of this researcher is that all students need to be given explicitly taught how to take tests on the computer, and given multiple opportunities to take practice tests on the computer. The QuizStar website was helpful for creating an online version of a test that contained images, but there were still several aspects of the program that were not user friendly.
for teachers or students. It was much more time consuming than it needed to be to enter
questions into the program, and the program would give students a final score based only on the
multiple choice questions. A student’s printed out test might show that he made a 100%, but that
score was only based on a limited number of questions. The teacher had to grade all of the other
answers, and the score might be much different than what it said at the top of the test.

Curriculum companies and the Smarter Balance Consortium need to develop programs
that will allow teachers and students to become familiar with the format of the tests. Just as
teachers model test taking strategies such as underlining answers in the text or writing out
calculations to the side, teachers need to be able to model strategies for computerized tests. The
difference of scores between the computerized and paper tests might have been much greater if
the class had not received direct instruction in strategies for taking computerized tests. More
research will need to be done over which type of questions are most effective on the computer
and what instruction is most beneficial for students preparing for a computerized test.
References


ED.gov. (n.d.). *No Child Left Behind: Elementary and Secondary Education Act (ESEA).*

Eyal, L. (2012). Digital Assessment Literacy -- the Core Role of the Teacher in a Digital


Aligning the Common Core State Standards with the Framework for 21st Century Skills.


Reading Score Differences at the Primary Grades. *Educational Assessment, 11*(2), 127-143. doi:10.1207/s15326977ea1102_3


doi:10.1080/00377996.2010.484445