# The Effects of Computer-Based Assessments on Student Scores for Spelling and Math

Kristi Krumperman

EDMA 5691: Educational Research

Dr. Jeremy Delamarter

Northwest University

7/22/2014

Table of Contents

List of Figures and Tables

Abstract

As technology continues to advance, its role in the modern classroom continues to change. With this change comes an increase in using computer-based assessments in the classroom. Little is known about the effect on student test scores when the students prepare with paper and pencil and then test on the computer. Which brings up the question: Does testing on the computer effect student test score? Using Senteo clickers and SmartBoard technology along with spellingcity.com, a group of fifth grade students were tested in spelling and math. It was determined that there was no significant difference in student test scores between paper and computer based assessments.

The Effects of Computer-Based Assessments on Student Scores for Spelling and Math

As technology continues to advance, its role in the modern classroom continues to change. With this change comes an increase in using computer-based assessments in the classroom. Since this is a relatively new concept, the education world at large is still studying the effects of computer-based assessment over paper-based assessment on academic performance. There are many districts in the state of Washington that have elected to have their yearly standardized tests administered online for the past four years and by the 2014-2015 school year, all schools in the state will be administering a new online assessment called the Smarter Balanced Assessment (Smarter Balanced Assessment).

Medina Elementary, the school at which I completed my student teaching, is in Medina, Washington, an affluent suburb of Bellevue. The median household income in 2012 was $160,000 (Medina, Washington). Additionally, with a large percentage of families working in the technology industry, all 27 of my students had at least one computer in the home. This is important because technology is a huge aspect and constant factor in the students' everyday lives. With this in mind, my research will seek to determine if there is a difference in test scores when students are prepared for the test with paper and pencil, but are then tested on the computer. In addition to this, I will be making note of the time that it takes to set up, administer, and grade tests that are taken both via paper-based and computer-based. If my research shows that there is a difference in student scores when they are tested on the computer, then this will give reason to incorporate more computer preparation to the curriculum. If the computer-based assessments yield lower test scores, then it is obvious that the students will need more

time to practice on the computer leading up to the high stakes testing. If the computer-based assessments yield higher test scores, then it is obvious that testing on the computers more often will have a positive impact on student learning. This leads to my research question, will testing on computers, in the traditional classroom, lead to a difference in scores in a fifth grade classroom at Medina Elementary when compared to a testing with paper and pencil?

## Literature Review

Currently, there are several studies that look into the effects of computer-based testing. Through these existing studies, trends, similar methodologies, and differing views have emerged. A trend that is common among the computer-based assessment literature is, simply, why the researchers are delving into this topic. Each article, in some form or another, stated that their purpose for doing this research was because of the higher prevalence of technology in the classroom and because of the increase of computer based assessments. Horkay, Bennett, Allen, Kaplan, & Yan (2006) stated that "in the 2005–2006 school year, for example, 22 states were reported to offer some type of online assessment" (p. 4). It can be assumed that this number has only increased. With the increasing use of computer-based assessment, researchers are curious how this impacts the students and their academic performance. Another common reason researchers have for pursuing this topic was the apparent time benefits of testing via computers. While it was not pursued in depth, many of the researchers discussed that using computer-based assessment has the potential to be beneficial because the "test results are scored instantly and reported in detail, their applications are flexible, effective and reliable…and enable the rater to easily adjust response time" (Yurdabakan, 2012, p.

177) and it "saves time grading, costs less than paper and pencil exams, automatically enters grades and potentially frees up class time" (Stowell & Bennett, 2010, p. 161). The Educational Testing Service (ETS) released an article in 2011 that detailed the possible advantages of implementing computer-based assessment. The ETS argues that computer-based assessments offer increased conveniences through self-proctoring, immediate scoring, integrated data management systems, flexible scheduling, diagnostic assessment and integration with instructional software, reach and speed, and preference. In addition, the ETS believes that computer-based assessment can lead to "a richer, more realistic experience that allows more direct measurement of the traits in question" (Educational Testing Service, 2011, p. 2). In today's classroom, every minute is valuable. If computer-based assessments can free up class and preparation time without having an adverse effect on student performance, then it is clear that a shift will occur in terms of testing.

In addition to their similar purposes for their research, there was also a major similarity in the results reported. While there were similarities the results were polarized. On one side, several studies reported that there was no significant difference in test scores between students that tested traditionally (pencil and paper) and students that had their tests administered on a computer. While there were some small differences reported, the general consensus has been that testing on the computer does not impact test scores, positively or negatively. On the other side, there were a number of studies that claimed that socio-economic status could impact academic performance, with low socio-economic students performing at a much lower rate on computer-based assessments than their wealthier peers.

Methodologies

In each of the quantitative studies that were reviewed, there was the inclusion of a control group and an experimental group. The qualitative studies used surveys and observations to reach their conclusions. Looking at solely the quantitative experimental designs, with the exception of the study on computer ownership, the participants received the same tests (regardless of which group they belonged to), but the mode of delivery differed. The control group had their tests administered via the traditional method while the experimental group had their test administered via the computer. The studies did differ in what specific aspect they were looking at. For example, one study was looking at anxiety while another study was testing the quality and the length of essays. I will be using a similar research design in my study.

Multiple Conclusions or Trends

Due to the fact that many of the studies were looking at different aspects of computer-based testing, there were some differences in the results. While the overarching result was that computer-based assessments did not have an impact on student scores (on studies that did not look at socio-economic status), there were some other, more specific, conclusions that were drawn.

In the study done on student attitude towards computer-based testing, it was determined that there was a "significant difference between the attitudes of private school students and state school students" (Yurdabakan, 2012, p.182) in regards to testing on the computer. Interestingly, it was concluded that the attitude of state school students was more positive in relation to computer-based assessment than the attitude of their private

school counterparts. In addition, this study also concluded that "computer-based applications increased the response time compared to paper-pencil ones" (Yurdabakan, 2012, p.178). These conclusions differ from the other conclusions because this study chose to focus on attitude of the student instead of their test scores.

In a different study, Stowell and Bennett (2010) hypothesized that computer-based testing would lead to lower levels of test anxiety than traditional testing. The results of their study did not support their hypothesis as they concluded that "students performed equally well under both exam conditions" (Stowell & Bennett, 2010, p.168). What is interesting about their conclusions is that the small sample of students that claimed to have an unusually high level of test anxiety when tested traditionally reported that "taking an online exam weakened the relationship between their test anxiety and poor exam performance, and reduced their test anxiety" (Stowell & Bennett, 2010, p.169). On the other hand, it was determined that the students that claimed to have low test anxiety when testing traditionally "online testing appears to promote test anxiety" (Stowell & Bennett, 2010, p.169). It was speculated by the researchers that this was due to "the added procedure of logging in to WebCT and relying on the technology to function properly, the extra responsibility of having to monitor themselves to avoid cheating or unfamiliarity with online testing" (Stowell & Bennett, 2010, p.169). In short, it was determined by the researchers that computer-based assessment was most likely to benefit the students that normally experience high levels of test anxiety because it removes them from the environment that triggers their anxiety (Stowell & Bennett, 2010, p.169). In relation to anxiety, the National Center on Educational Outcomes (NCEO) published an article that looked at which states were using computer-based assessment

programs for the standardized tests and which states were not. The NCEO also looked into whether or not these states were taking students with disabilities into account. It was discovered that many states were not looking into the effect of computer-based assessments on students with disabilities. The NCEO, in their conclusion, urged the education community to expose students with disabilities to testing on computers prior to the standardized tests claiming that "without previous experiences and facility in using computers, it is possible that the introduction of the computer itself could depress the performance of students who have not previously used a computer for instruction" (Thurlow, Lazarus, & Albus, 2010 p.12). Knowing that the term "disabilities" covers a wide spectrum of diagnoses, this could have serious implications on a significant percentage of the population.

In yet another study, conducted by Fairlie and Robinson, the effects of computer ownership on test scores was examined. After placing over 1,000 computers in homes that did not already have a computer, the researchers found "no evidence that home computers had an effect (either positive or negative) on any educational outcomes, including grades, standardized test scores, or a host of other outcomes" (Fairlie & Robinson, 2013, p.1). If computer ownership does not does not have an impact on the grades or test scores of low income children, then it opens the doors for more computer-based testing to be added to the curriculum. On the opposite end of the spectrum are two studies that examined the effect of the digital divide on academic performance. The digital divide can be defined as "the gap between the students who have access to digital technology at home and those who do not" (Sun and Metros, 2011, p.154). This divide includes computers and/or internet access. There are several contributing factors to the

digital divide with socio-economic status being the most likely contribution to the divide. Gunduz (2010) lead a study that looked at 375 elementary schools in Turkey. This study was seeking to determine if there was a "parallel relation between the increase of average grade of students and the increase of having computer and internet connection in their home" (p.43). Gunduz (2010), through a survey method, determined that, out of 375 schools, "most of the students, who ranked in top ten at their own class levels…have computer and internet in their homes, and those ranked in 'last ten' do not have computer and internet at home" (p.50). The implications of these findings are huge and they directly contradict the results of the study done on the effects of computer ownership. This difference in results could be explained by Eamon (2004). In his publication, he cited that it was determined that

> a higher percentage of low-income youth used the computer to play video games daily, compared to their wealthier peers….because their parents are less able to provide educational software, computer hardware, technical assistance, and supervision, compared to wealthier parents. (p. 95)

In summary, low income students are less likely to use a computer in the home for educational purposes when compared to their wealthier counterparts. This alone could have an impact on academic performance. In the digital divide study, done by Sun and Metros (2011), it was determined that "socio-economic status and social capital not only affect students' access to institutional resources, but also affect their opportunities to use technology" (p.159). Sun and Metros (2011) believe that when used properly, technology can be an effective educational tool, but that it might not be an effective educational tool for students from low socio-economic status unless they are first taught

how to access and use the computers for educational purposes. From these four studies, a variety of conclusions can be drawn. It can be assumed, that computer ownership does impact student learning, but just giving computers to students that are of a lower socio-economic background is not going to solve the problem. There are issues at hand that are outside of the realm of education, but a first step could be teaching students how to use computers to advance their education.

In addition, there was a study that looked at the effects of computer-based assessment on GED testing scores. In accordance with numerous other studies, it was determined that there was no significant difference in test scores, but it was also determined that computer-based testing made it "easier to accommodate to specific special needs" (Higgins, Patterson, Bozman, & Katz, 2010, p.7). These findings could have interesting implications when you look at the time that goes into teachers having to make so many testing accommodations, all of which cut into preparation and instruction time.

The study conducted by the National Assessment of Educational Progress chose to focus on the quality and length of written essays done via paper and pencil and on a computer. While it was determined that there was no significant difference in quality, it was discovered through this study that "students who used computers when learning to write not only produced written work that was of higher quality and greater length, but were more engaged and motivated in their writing" (Horkay et al., 2006, p.36). With this information, the researchers speculated that practicing writing on the computer could positively impact the students' future writing on the computer as well as their writing on paper. On the other hand, they also speculated that "students with little practice writing

on computer will not score as highly in an online writing test....not necessarily because [they] are less skilled writers, but because they are less skilled writers on computer" (Horkay et al., 2006, p. 36). If this is true, then it could be said that students, in order to have the best chance at passing their standardized tests, should have ample opportunities to practice their writing on the computer.

## Gaps in the Research

While these articles do a fairly good job at covering different aspects of the effects of computer-based testing, I found that there are some facets that are not mentioned. This led me to be interested in seeing if there is a difference in time spent on computer-based testing versus traditional testing. While it is obvious that computer-based testing takes less time to grade, I became curious to see if computer-based testing took longer to set up or longer to administer than traditional tests. I also became interested in what the effect of moving away from paper-based assignments and assessments will have on the quality of writing, communication, and work in multiple areas of study, but am aware that this focus would require me to run a study over an extended period of time.

## Methodology

For this research project, I performed a quantitative study. I used a quasi-experimental research design that included a pre-test and a post-test. This study included a control group and an experimental group. By having a pre-test and a post-test, this allowed me to compare the results to determine if there was an effect of computer-based assessment on student scores, either negative or positive. I also looked at the amount of time that it took me, as the teacher, to set up, administer, and grade the tests. The

purpose of this is to determine if there is a difference in class and preparation time spent between the two testing methods.

Sample

This research is a quasi-experimental design because my sample is pre-determined. I used the students from the 5th grade class that I did my student teaching in. In this class, there were 27 students (24 of the students participated in the math study and 20 of the students participated in the spelling study). Due to the demographics of the class, I used stratified sampling to choose the groups. There were 8 students that received gifted/enrichment services. They were assigned at random to either the control group or the experimental group. There were 2 students with documented learning disabilities in this 5th grade classroom. Both of these students have IEPs for reading and were randomly assigned to either the control group or the experiment group. Using stratified sampling helped me to ensure that the students with learning disabilities and the students that receive enrichment/gifted services were not over or under represented in either group. This helped to control for confounding variables that might have had an effect on the outcome of the experiment. It also allowed me to more accurately analyze the effects of the computer-based assessments on students with documented disabilities.

Procedures

The first step was determining the subjects I wanted to test. I chose to test mathematics and spelling. After selecting the subjects, I wrote a pre-test for each subject that was administered to the entire class via the same method, paper and pencil. After I got the results from the pre-test, I developed a lesson for each subject and randomly assigned the students to either the experimental or control group through the stratified

sampling method. Next, I taught my lessons based upon the information that I gathered from the pretests. For the spelling lesson, we focused on the language functions and the sounds that they make. The students were also asked to sort the words based on similar sounds and spelling. For the math test, the students were taught what a benchmark fraction was and how to use them to estimate when adding and subtracting fractions. The students all received the lesson in the same manner and were not split up based upon whether they were in the control or experiment group. In order to control for accurate results, the post tests had the same questions and words that the pretests had. For the spelling test, the students in the control group took the test on paper while the experiment group took their test online at spellingcity.com. For the math test, the students in the control group took the test on paper while the experiment group took the test using the SmartBoard and Senteo Clickers. The students in both the control and the experiment group were given ample scratch paper. The students all took the test at the same time and were held to the same testing standards. These standards include, but are not limited to, voices must remain silent unless they are asking the teacher a question, they may not look at their notes, and they may not look at their neighbor's test. I stood so that I could see all of the computer screens. This eliminated the possibility that the students accessed the internet or any notes while taking their test. During the tests, the amount of time that it took the students to finish was recorded. After the tests were completed, they were graded by group. This allowed me to record the time that it takes to grade each method of testing.

## Results

The first experiment was the spelling experiment. The following figures and tables show how the students scored on the pretest and the posttest. The student scores are divided between the control group and the experiment group. The pretest and the posttest were identical in content and the highest score was 20.

Figure 1

*Pre and Post Test Scores for Control Group for Spelling*

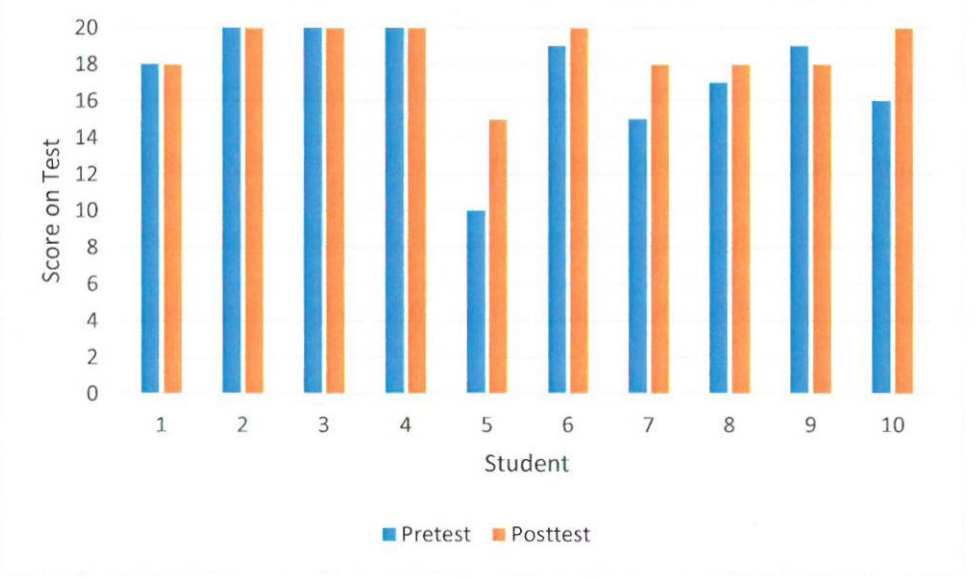

Figure 2
*Pre and Post Test Scores for Experiment Group for Spelling*

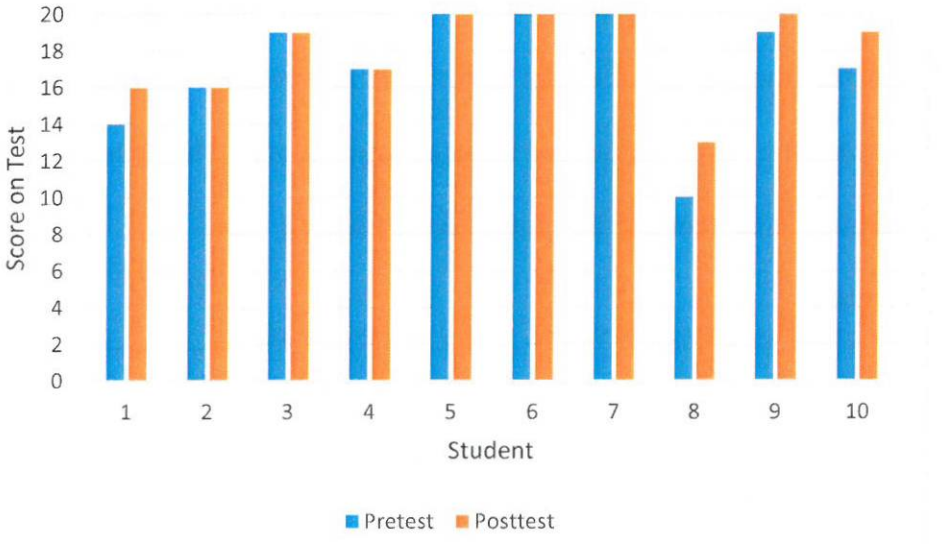


Table 1
*Pre and Post Test Averages and Stastical Significance for Spelling Test*

| | N | Pre-Test Average | Post-Test Average | Sig. (2-tailed) |
|---|---|---|---|---|
| Control | 10 | 17.4 | 18.7 | 0.07 |
| Experiment | 10 | 17.2 | 18 | 0.053 |

The data for the control group has an average pretest score of 17.4 and an average posttest score of 18.7. The data for the experiment group has an average pretest score of 17.2 and an average posttest score of 18. In the control group, 5 students improved their test score and 1 student lowered their test score. In the experiment group 4 students improved their test score and no students lowered their test score. The change in test scores was not significantly different for the control group as I was looking for a number under 0.05 to determine that my experiment design had an impact on the student test scores. The experiment group came close to significance at 0.053, but with the average

test score only increasing by .8 in a sample size of 10, it cannot be said with certainty that testing on the computers has responsible for this increase in scores.

The second experiment was the math experiment. The following figures show how the students scored on the pretest and the posttest. The student scores are divided between the control group and the experiment group. The pretest and the posttest were identical in content and the highest score possible was 5.

Figure 3

*Pre and Post Test Scores for Control Group for Math*

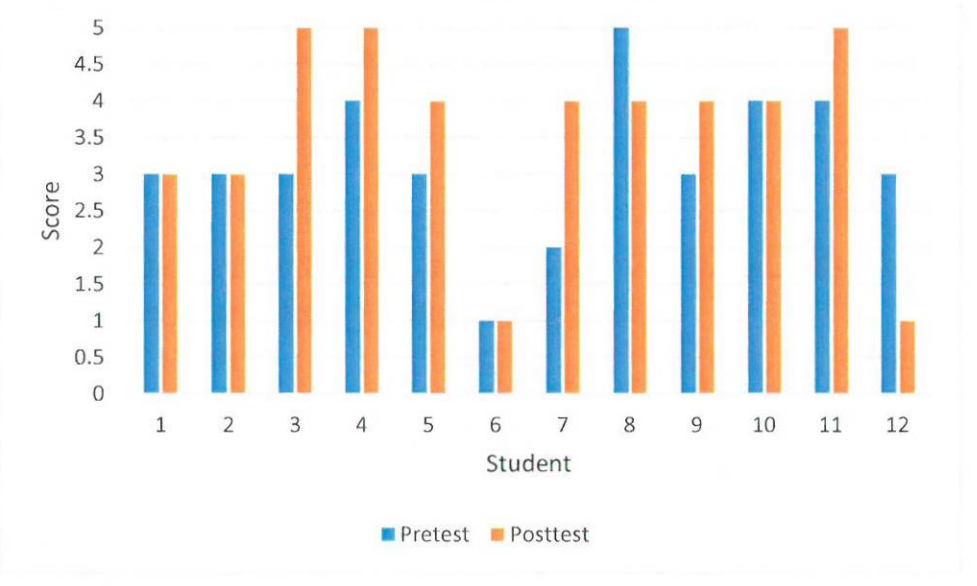

Figure 4

*Pre and Post Test Scores for Experiment Group for Math*

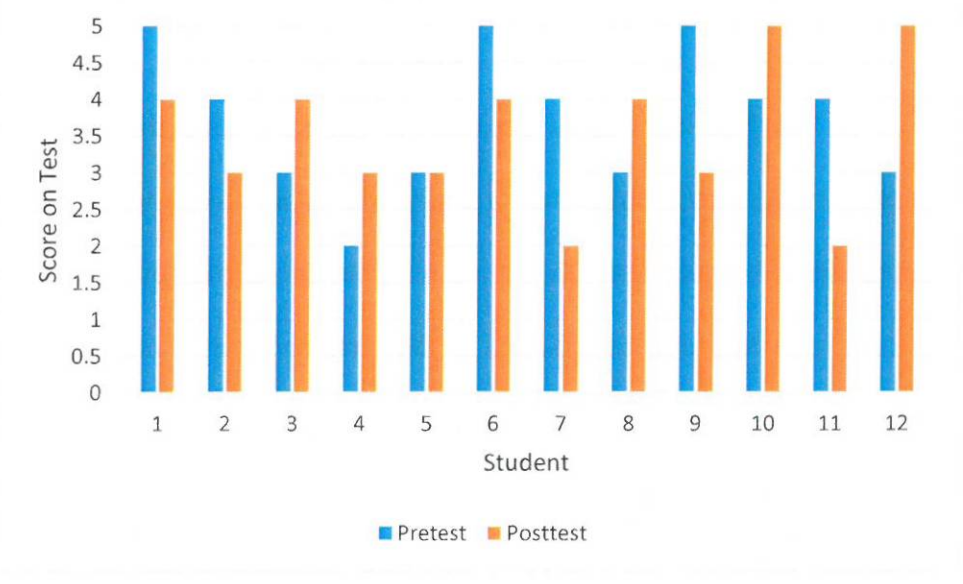


Table 2

*Pre and Post Test Averages and Statistical Significance for Math Experiment*

| | N | Pre-Test Average | Post-Test Average | Sig. (2-tailed) |
|---|---|---|---|---|
| Control | 12 | 3.8 | 4.3 | 0.241 |
| Experiment | 12 | 4.5 | 4.2 | 0.555 |

The data for the control group has an average pretest score of 3.8 and an average posttest score of 4.3. The data for the experiment group has an average pretest score of 4.5 and an average posttest score of 4.2. In the control group, 6 students improved their test score and 2 students lowered their test score. In the experiment group 5 students improved their test score and 6 students lowered their test score. Unlike the spelling test, which was reaching significance, the math test scores indicated that there was no significant difference in student test scores from pre-test to post-test in either the experiment group or the control group.

The third experiment conducted was on the amount of time that it took to prepare, administer, and score the math and spelling tests. The combined amount of time for the

two control groups was 1 hour. The combined amount of time for the two experiment groups was 35 minutes.

## Discussion

While there is a slight difference between students in the control group and students in the experiment group for each experiment, there does not appear to be a significant difference in student test scores when the student prepares on paper and tests on the computer or the SmartBoard. In the spelling experiment, the control group saw a 1.3 point increase in the average student test score. The experiment group saw a .8 point increase in the average student test score. The control group scored, on average, 3.5% higher than the experiment group. In the math experiment, the control group saw a .5 point increase while the experiment group saw a .3 point decrease. The control group scored 6% higher than the experiment group on the posttest.

The most interesting part of the experiment was the decrease in students' scores from the math experiment group. 50% of the experiment group experienced a decrease in test score when they tested using the Senteo clickers. It is possible that this is due to the students' lack of working out the problems. Every student from the control group showed their work for each problem. The students from the experiment group were all given scratch paper, but not a single student showed their work for every problem and 4 of the 12 students did not show their work for any of the problems. This is a potential area of concern because showing work allows students to check their work for accuracy. Furthermore, it is possible that the students' scores declined because of the use of the clickers. When filling out a multiple choice test on paper, a student is much less likely to

accidentally circle the wrong answer. When a student is simply selecting a button, there is a greater chance of error in selection.

The results from this experiment agree with the literature for this topic; there is no significant difference in student scores between the experimental and the control group. It is possible that this was due to the high socio-economic demographic of the selected school (under the assumption that these students have a higher access to technology because of this status). In the future, students will be required to solve the problems on paper when testing on the computer. This might prevent students from rushing or committing a small mathematical error.

In regards to the time spent on the tests, there is a significant difference. This is directly related to the fact that I did not have to grade the computer tests or run copies. For both tests, reports were emailed to me with all of the students' scores. This is an important factor because with teachers being stretched for time, being able to eliminate almost 50% of the time needed for a test could allow teachers to spend more time interacting with the students, preparing for projects, grading other assignments, or assisting struggling students. Given that there is not a significant impact on student test scores when testing on the computer or the SmartBoard, I plan to implement more computer based tests. This will, not only, decrease the amount of time I spend preparing and grading the tests, but will help the students prepare for the standardized tests which are administered via computer.

## Recommendations

There are a few threats to the validity of this research. Primarily, the sample for this study was not random, as I used a predetermined set of students. Another threat to

validity was the potential that any students experienced test anxiety that was heightened by either of the testing methods. This could have skewed the data. In addition, it is possible that student attitude may be a threat to validity based on whether students enjoy testing on computers or not.

Moving forward, I would like to look deeper into the cause of the decline in test scores. While it can be assumed that this was a result of students not showing their work, I would like to eliminate any other possible reasons why 50% of the students experienced a decline in test scores when they tested on the SmartBoard. Additionally, I would like to do a qualitative study on student test anxiety in relation to computer-based testing. I would like to compare anxiety levels with test scores. In a study done by Fritz and Marszalek (2010) it was determined that students who tested on computers reported less anxiety than the students who tested on paper and pencil. This study, however, compared two different schools that tested at different frequencies. The students who tested on computers were tested more frequently. I would like to replicate this study, but alter the testing conditions so that each group of students was being tested the same amount of times. Furthermore, I would like to look into the quality of writing from students when responding on paper versus on the computer. I would also like to conduct all of these experiments at different schools with different resources and different socio-economic statuses. According to a study done by Lebens, Graff, and Mayer (2009) students of a lower socio economic status "express feelings of threat and tension when they think of using a computer" (Lebens, Graff, and Mayer, 2009, p. 263). This study was done in a district with high access to technology and technology classes. In a district like Bellevue which has a wide variety of socio-economic background, it would be interesting to do a

district-wide study on student perception of technology compared to the demographics of the district. This is important considering the STEM (Science, Technology, Engineering, and Mathematics) Initiative that multiple elementary schools in the Bellevue School District are participating in. These schools include students from a wide-variety of socio-economic backgrounds. The goal of the initiative it so get students excited and interested in STEM careers (Elementary STEM Program). With Lebens, Graff, and Mayer's study in mind, the STEM initiative could have interesting and mixed results.

References

Eamon, M. K. (2004). Digital divide in computer access and use between poor and non-poor youth. *Journal of Sociology and Social Welfare* 31(2), 91-112.

Educational Testing Service (2011). Practical considerations computer-based testing. *Educational Testing Service.*

Elementary STEM Program. (2013, January 1). *Bellevue School District.* Retrieved July 3, 2014, from https://www.bsd405.org/news-events/news/2013-school-year/elementary-stem-program.aspx

Fairlie, Robert W & Robinson, Johnathan (2013). Experimental evidence on the effects of home computers on academic achievement among schoolchildren. National poverty center working paper series #13-02. *National Poverty Center, University of Michigan.*

Fritts, Barbara E & Marszalek, Jacob M (2010). Computerized adaptive testing, anxiety levels, and gender differences. *Social Psychology of Education: An International Journal* 13(3), 441-458.

Gunduz, Hasan Basri (2010). Digital divide in Turkish primary schools: Sakarya sample. *Turkish Online Journal of Educational Technology* 9(1), 43-53.

Higgins, Jennifer; Patterson, Margaret Becker; Bozman, Martha; & Katz, Michael (2010). Examining the feasibility and effect of transitioning GED tests to computer. *Journal of Technology, Learning, and Assessment* 10(2).

Horkay, Nancy; Bennett, Randy Elliot; Allen, Nancy; Kaplan, Bruce; & Yan Fred (2006). Does it matter if I take my writing test on computer? An empirical

study of mode effects in NAEP. *Journal of Technology, Learning, and Assessment* 5(2).

Lebens, M; Graff, M; & Mayer, P (2009). Access, attitudes and the digital divide: children's attitudes towards computers in a technology-rich environment. *Educational Media International,* 46(3), 255-266.

Medina, Washington. (2012, January 1). *(WA 98039) profile: population, maps, real estate, averages, homes, statistics, relocation, travel, jobs, hospitals, schools, crime, moving, houses, news, sex offenders*. Retrieved July 8, 2014, from http://www.city-data.com/city/Medina-Washington.html

Smarter Balanced Assessments. (n.d.). *Smarter Balanced Assessment Consortium*. Retrieved July 6, 2014, from http://www.smarterbalanced.org/smarter-balanced-assessments/

Stowell, Jeffrey R & Bennett, Dan (2010). Effects of online testing on student exam performance and test anxiety. *Journal of Educational Computing Research* 42(2), 161-171.

Sun, Jerry Chih-Yuan & Metros, Susan E. (2011). The digital divide and its impact on academic performance. *US-China Education Review* 12(6), 157-177.

Thurlow, Martha; Lazarus, Sheryl S; & Albus, Debra (2010). Computer-based testing: Practices and considerations. Synthesis Report 78. *National Center on Educational Outcomes.*

Yurdabakan, Irfan (2012). Primary school students' attitudes towards computer based testing and assessment in turkey. *Turkish Online Journal of Distance Education,* 13(3), 177-188.