**Declining Test Scores: Paper and Pencil Versus Online Reading Comprehension Assessments**

Patrick Walker, David Vasilia, Sam Salas, Kyle Klimenko and Rafael Ruvalcaba

EDUC 607

Dr. Murillo

Spring 2014

**Abstract**

The purpose of this study is to test students and measure results that will indicate if paper or digital versions of text will result in varying performance outcomes. The reality of educational institutions moving towards online and digital forms of gathering information, completing and turning in assignments, and taking assessments is becoming more present in our classrooms. In this study we measure and report on the differences between paper/pencil, and online reading comprehension assessments. This quantitative study of 50 California third graders.

**Table of Contents**

**Introduction……………………………………...…………………………………………..4**

**Review of Related Literature……………………………………………………………………………………… 5**

**Assumptions………………………………………………………………………………… 20**

**Anticipated Results/Foreshadowed Problems………………………………………… 20**

**Definition of Terms…………………………………………………………………………. 21**

**Significance of the Proposed Study…………………………………………………….. 22**

**Design and Methodology………………………………………………………………….. 23**

**Instrumentation/Data Collection…………………………………………………………. 24**

**Data Treatment Procedures……………………………………………………………….. 25**

**Presentation of Findings…………………………………………………………………… 26**

**Limitations of the Design………………………………………………………………….. 31**

**Conclusion…………………………………………………………………………………… 31**

**Recommendations for Further Research……………………………………………….. 32**

**References....................................................................................................................... 33**

**Introduction**

The digitization of music, images, video, documents, and other forms of media has facilitated the work in many fields. Communication has been quickened by email, movies are restored by creating digital versions, music is digitally remastered to improve quality, and literature is made readily available through online databases. Historically, education is always just behind the modernization of practices and demonstrates that in its late introduction to the digitization of practices in public schools. All testing is now becoming digital and students are now being required to learn a new set of skills to accomplish academic tasks. Districts and teachers can now attain student results quickly, analyze data, and improve their practices.

Measuring student reading comprehension is formally done by an assessment administered by a classroom teacher. The assessments usually contain a fluency score along with a comprehension score that allows the teacher to find a reading level. Students are then grouped by levels and teachers differentiate in small groups to target instruction. Traditionally, these assessments are done on paper to allow teachers to closely monitor the subtle errors or miscues of student reading. Reading assessments come with comprehension questions that increase in difficulty as the student goes up in reading levels. Questions revolve around the students’ ability to make sense of the reading. They can ask students to categorize and classify, note details, find the main idea and supporting details, know what viewpoint, make inferences, or use any other strategy that helps the student understand what they have read.

California school districts are now piloting Smarter Balanced assessments. These assessments are web-based assessments that require students to complete a performance-based task on a digital device. This new generation of digital natives welcomes these practices at school since most use a digital device outside of school. However, much can be said about the effectiveness of digital versions of reading material. Some argue that digital reading material may not be the best medium to teach or assess students with. The reasons vary, but this study focuses specifically on reading comprehension results. What medium is best suitable to assess reading comprehension at the primary grades? Will traditional print, or a digital reading content and assessments render better results?

Overall, this paper seeks to answer these questions. We look at third grade students who have taken both pencil/paper, and digital reading comprehension assessments. The goals and purpose of our study is to determine if there is a medium in which students perform better. In this study, we analyze the results of reading comprehension scores using digital and print forms, and report the difference.

**Literature Review**

A body of research relates to our topic of reading comprehension and digital text. This research is both quantitative, and qualitative. It spans a broad spectrum of grades from the elementary level, to the university. In the following pages we outline and summarize ten studies. These studies contributed significantly to our work, and assisted us in determining the current paradigm.

***Digital Reader vs Print Media***

Gill and the authors of “Digital reader vs Print Media” performed a quasi-experimental quantitative study were all subjects (not random) receive interventions to test differences in group means for the independent variables. The purpose of the study was to compare user satisfaction, reading accuracy and reading speed between standard paper print, Sony eReader and Apple iPad for patients with stable wet age-related macular degeneration (AMD). It is already known that patients with AMD have difficulty reading texts and that digital devices may provide better magnification of text, but it is unclear if digital devices are preferred or if they reduce reading speed. This study showed that patients with low vision and AMD can benefit from using digital devices for reading.

The 27 patients used in the study had stable wet AMD, in one or both eyes ranging from 20/25 to CF. Patient average age was 78.9 years old. Standardized texts (New Courier) were validated through the Hahn study for reading speeds and text contents were found. Then text sizes were chosen (size 12, 16, 24, 32, 50, and 80) and font sizes were calibrated on printed paper and both digital readers using a microcaliper. All reading assessments were completed in the same room to ensure zero difference in room luminance.

Patients started with the smallest print size they could comfortably read on paper. They read aloud and words per minute were recorded by a research assistant. Subjects were asked to read aloud as fast as they could without making mistakes and without correcting mistakes. Incorrectly read words were subtracted from the word count. Next, subjects were instructed on the use of digital readers including how to scroll through text. Each patient was then asked to read texts aloud on each of the digital reader while reading speed was recorded as before. At the conclusion of the readings, subjects completed an assessment on ease of use and clarity of print for the device after each reading.

The data recorded for analysis included patient font size selected, oral reading speed, and a survey assessing their ease of use and clarity of print for paper print and the digital devices. The data was very straight forward. Data was first categorized into groups based on preferred text size, and then the mean was determined for each group to determine average words read per minute. Patient preference data was combined to determine the mean of all subjects, not groups.

Font size selection divided the patients into three groups (9 subjects in each), 12-16 size font, 24 size text font, and 32 size font or greater. All groups showed the fastest oral reading speed using the Apple iPad, followed by paper print and leaving the Sony eReader with the lowest number of words per minute. Total averages from all groups are as follows:

· Apple iPad – 117.1 WPM

· Paper print – 113.2 WPM

· Sony eReader – 110.6 WPM

Additionally, patients determined the paper print had the highest “ease of use”, followed by the iPad. However, patients rated the paper print below the iPad when determining “clarity of text”.

The authors concluded that digital devices may have a use in assisting low-vision patients as they were able to read more quickly on the Apple iPad. Though the patients preferred the paper print over the digital devices, this may be attributed to patient ages and their lack of exposure to technology. It was also determined that the iPad had the best “clarity of text” because of the high contrast between the lettering and background.

It is believed that the Sony eReader had the least success in the study because the monochromatic gray scale of the device had much less contrast between the text and background than either paper print or the Apple iPad.

The author cautions that the patients were all tested using both eyes to read, while some of them only had AMD in one eye. Though this concern is valid, I think that the results are still applicable to patients with low vision, but text sizes and reading speeds may be slower in patients who have AMD in both eyes.

Additionally, the authors point out that it is likely that subject age significantly effects their preference to use paper print media as they have had much less exposure to digital devices than someone half their age would.

***It's Time to Turn the Digital Page: Preservice Teachers Explore E-Book Reading***

Lotta C. Larson has been an assistant professor at Kansas State University for 5 years prior to the study being published. Before her work at Kansas State University, she worked as a 5th and 6th grade classroom teacher. For the study, he used students in two sections of her university methods course entitled “Teaching Literacy in Grades 3-6” as her subjects.

As a teacher educator, the author believes it is her duty to give teachers the support and skills needed to be successful at integrating literacy using new technologies. Larson has three objectives with this lesson. The first is to help her subjects “gain firsthand experience in reading an e-book”. Next, she is trying to have her students “consider text factors and reader factors that support comprehension.” Finally, she is hoping to teach her students how to integrate e-books into their future classroom lessons.

The study took place at Kansas State University and included 49 of the author’s literacy methods course. Participants included 3 males and 46 females, all of whom were preservice teachers. As a course requirement, participants read the children’s book *Moon Over Manifest* by Clare Vanderpool and completed both a prereading questionnaire and a postreading questionnaire. 47 participants had their own computers or e-reading devices to complete the assignment, 2 students borrowed Amazon Kindles from the author.

The author was a complete insider in the study as she assigned the work to her current students and lead all classroom discussions. The study began as students filled out a written prereading questionnaire that includes 5 open ended questions. Over the next three weeks they read the book *Moon Over Manifest* by Clare Vanderpool which included 9 class sessions. During class sessions, the following occurred: (a) Discussed the story and e-book reading experience in small groups. (b) Discussed the story and e-book reading experience in whole class conversations lead by the author. (c) Shared lesson plans and suggestions on how to integrate e-books into the classroom. At the end of the three week reading period, a postreading questionnaire that includes 9 open ended questions was completed in writing and submitted to the author.

Data was collected as subjects completed a prereading and post reading questionnaire in writing at the beginning and end of the three week reading period. Additional data sources included “field notes, lesson plans, and audio recorded classroom discussions”, along with students digital e-book notes and bookmarks.

The researcher combined the comments from the prereading and postreading questionnaire to look for patterns that emerged. Completing the activity did give preservice teachers firsthand experience in reading an e-book, and just over half (26 of 49) of the subjects shared that reading with an e-book did support their reading comprehension when compared to reading a traditional book. While 15 subjects reported that digital reading neither hindered nor supported their experience.

All subjects made use of at least one of the e-book features (ex: highlight, notes, bookmarks, font size) most taking advantage of many features showed that text factors and reader factors were considered.

Results were hardly discussed when it came to integrating e-books into the classroom. One subject response is pointed out discussing a balance is needed between traditional books and e-books. With a concern that computers and other devices may not be easily available at all schools and homes.

***Research on Learners’ Preferences for Reading from a Printed Text or from a Computer Screen***

The article, *Research on Learners’ Preferences for Reading from a Printed Text or from a Computer Screen*, reports on the research conduced by Carrie Spencer. Spencer is an instructional designer and the Director of the Centre for Educational Technologies at Royal Roads University. She has interests in I.T. and the role that technology plays in communication. When the study was written, Spencer was part of a four person project management team leading in the development of major IT undertakings at the University.

This article is primarily a qualitative study focusing on the methods of reading, factors that affect reading, and implications for future distribution of course materials at Royal Roads’ University. However, the author of the study also includes some quantitative measures and descriptive statistics. Overall though, the study focused on quality and utilized individual interviews in addition to Likert-type scale questions.

The purpose of Spencer’s study is to determine learner’s preference for reading course materials in the Royal Roads University School of Business’ graduate and undergraduate degree programs. The university offers a hybrid program for its students where they must take some courses online, and others in short residency periods. Spencer writes in her rationale that “Because most RRU distributed learning courses are offered via the internet, university administrators questioned the need to incur the shipping costs necessary to provide learners with the printed course materials. The focus questions for the study are as follows:

1. By what method (online or print) are text-based course notes most often read by learners currently participating in School of Business programs at RRU?

2. How do factors such as age, eyesight, and comfort with computer technology affect learners’ preferences for course reading material?

3. What implications do these findings have for the most effective distribution of course materials to RRU learners?

Spencer’s study took place over the Internet and through follow-up phone conversations. The participants in the study were 500 graduate and undergraduate students. Of those surveyed, 254 students responded to the survey. Six respondents were randomly chosen to participate in follow-up telephone interviews. Salient characteristics include an approximate 50% of respondents being female, and about half of all graduate and undergraduates in the MBA and BCom programs responding.

Data for the study were collected in two ways. First, Spencer emailed surveys to 500 graduate and undergraduate students. These students were enrolled in Royal Roads’ University’s hybrid MBA and BCom programs. Secondly, Spencer randomly chose (every third participant) learners for follow-up telephone interviews.

Spencer informs the reader on page 39 that the study was an emergent design. Analysis of the survey data does not seem to surpass calculation of the mean. These means were then changed into percentages of the whole. These descriptive statistics were used to support the quotations from the follow-up phone interviews. It was not made clear in the study how the analysis of phone interview data were analyzed.

In general Spencer’s research supported the paradigm that people prefer reading paper text to screen text. She writes, “The convenience of paper for reasons of portability, reliability, annotation, highlighting and ergonomics consistently made it the preferred form for printed text.” Many of the participants in the study in the study read their course materials on public transportation, planes, at the gym, or waiting to pick up their children. Most of these areas do not allow for easy access to Internet connections.

***The Effects of Reading from the Screen on the Reading Motivation Levels of Elementary 5th Graders***

Aydemir and Ozturk’s article, *The Effects of Reading from the Screen on the Reading Motivation Levels of Elementary 5th Graders*, utilized a randomized control group pre-test/post-test model for their research. This is a true experiment in that the groups were randomized with fifty percent of participants in a control group, and fifty percent of participants in an experimental group.

The general purpose of the study was to identify the effects of texts read from the screen on the motivation levels of Turkish 5th graders. The researches sought to answer one question in their study: Is there a difference between the reading motivation levels in the control and experimental group? The reader can infer beyond the authors’ stated question to ask, Is there a difference in motivation between those who read on a screen, and those who read from printed text?

Answering their research question, and the one that we may infer, can help us to better understand the implication of transitioning our state assessment systems to an all on-line format. Currently, California is part of the Smarter Balanced Assessment Consortium (SBAC). These high-stakes assessments are only offered in an online format. The motivation levels of our students participating in these assessments are sure to affect outcomes.

The study took place in an elementary school in Sakarya Turkey during the second term of the 2011-2012 school year. The participants were sixth 5th grade students. The school had continuous Internet access, and the authors state that measures were in place to ensure students in the study had basic computer literacy.

The date recorded and used for analysis in this study were student response on the Reading Motivation Scale for Text. This scale was developed by the authors of the study. Data was collected from the pre and post tests of all sixty 5th graders.

The authors of the study utilized causal data analysis techniques. The students were randomly assigned to either a control or experiment group. All students were given pre and post-test measures. The data collected from the Reading Motivation Scale for Text was statistically analyzed to determine the mean, and standard deviation in, and between groups. Further statistical analysis was conducted using a two-factor analysis of variance (ANOVA). Subdimensions were studied to explore if they had changed significantly between the pre and posttests.

Overall, the findings of the study are that reading from the screen causes a decline in students’ reading motivation. However, the mean motivation levels of all students declined between the pre and posttests. The motivation levels of the control group actually decreased more than those students reading from the screen.

Aydemir and Ozturk conclude that reading from a screen did decrease the motivation levels of Turkish 5th graders. They also note that when the experimental group was compared to the control group, “the decline of the experimental group was less.” They conclude that this could be attributed to the experiment being an 8-week trial. Furthermore, the participants, 5th graders, “get bored when faced with activities that do not interest them, and lose their interest in 10 minutes.” Also concluded were recommendations for further study. Future researchers might choose to include poetry or other genres, shorten the length of the study, include animations, or choose another grade level.

The cautions of the study were substantial. First, the sample size, and location of the study make the results not entirely generalizable to larger populations. Also, the mixed findings from the study make it hard to determine the practicality on a day-to-day basis. There are so many factors that may have played into the decrease in the students’ reading motivations that it is hard to decipher if the experimental variable was truly measured. It would have been interesting to retest six weeks after the posttest, and check if motivation levels had rebounded.

***Overcoming screen inferiority in learning and calibration. Computers in Human Behavior***

Taking into account the individual differences in beliefs regarding the effectiveness of learning on screen versus on paper, by extending a line of research analyzing reading comprehension that can be performed comparably on both media, Lauterman and Ackerman found that peoples’ medium preference affected their metacognitive processes when learning from texts (Ackerman & Lauterman, 2012). They go on to explain that continuous text is widespread in computerized learning environments, and in many cases these environments offer no special features that are not found on paper. For instance, digital media provide an assortment of on-demand textual information for developing professional competence, like providing access to user reference books or academic papers.

Students face computerized reading comprehension tasks in their studies, and higher education candidates face them in online screening exams like the G.M.A.T. (Graduate Management Admission Test), thus it is important to consider the ways self-regulated learning is affected by the medium on which one learns (Ackerman & Lauterman, 2014), and because several studies have found screen inferiority in subjective and objective learning, the goal of their study is to offer methods of overcoming screen inferiority. Performing the studies while considering the study medium, screen versus paper, and participants medium preference as factors in the improvement.

The experiment included informing participants in advance that the test questions would examine both memory for details and higher-order comprehension. Each participant studied and was tested on two texts in each condition. Eighty-seven undergraduate engineering students were the subjects with a ratio of half female, half male. They were split into groups of eight participants in a small computer lab each group working on the same medium. They were told that they would be studying for a multiple choice test that would assess both their memory for details and their higher-order comprehension.

As a prediction of performance, the participants were asked to drag an arrow along a continuous 25%-100% scale to indicate how well they thought they would do on test questions that involved (1) memory for details and (2) higher order comprehension.

The results of experiment 1: Where the participants were randomly selected and not given an option as to which type of test they were going to be given (screen or paper) screen inferiority was verified where the participants placed confidence of performance (on average) at approximately 75% for their computer test while, scoring (on average) only in the high 50% range. Paper results proved to be somewhat similar in confidence at mid 70% while actually scoring in the mid 60% range.

The results of experiment 2: When the participants were given a choice of preference (screen or paper) and given the results of their practice tests, the results yielded significant improvements in perceived outcome and actual outcome, virtually eliminating screen inferiority and attenuation of overconfidence.

***Reading linear texts on paper versus computer screen: Effects on reading comprehension***

Mangen, Walgermo, and Bronnick’s article about “*Reading linear texts on paper versus computer screen: Effects on reading comprehension*” points out the ongoing transition of reading from print to screen, it also points out that as of 2012 (the time the article was published), much of the implications of this transition from paper to computer screen reading on comprehension.

Previously referenced studies have shown an increase in cognitive load associate with “hypertexts”, and points out that not all text is hypertext and therefore asserts that not all digital reading is subject to the increase in cognitive load associated with hypertext.

Mangen, Walgermo, and Bronnick’s study was implemented in more of a natural setting (classroom) in a “test-like situation” and extends previous research within this field with regard to ecological validity. The stimulus material consisted of two authentic texts that are both quite representative for their respective genres (narrative and expository). Effort was made to “be as close to identical as possible across presentation media.”

Mangen, Walgermo, and Bronnick’s hypothesized (1) that they would find, as in previous research, better reading comprehension when text were read on paper, and (2) that the expository text would be more affected by reading modality than narrative, due to the possible higher cognitive load introduced by the top of the text. The additional motivation for this study is that the scoring system which will be employed in Norway will have direct influence using this study for policymaking and test development.

The experiment in Mangen, Walgermo, and Bronnick’s article describes two groups, one read texts digitally/on screen, while the other read texts on paper. Both groups answered questions digitally/on screen. The digital reading was done on 15” LCD monitors using Adobe Reader version 9.4. For group 2 the same texts were printed on A-4 size paper. The was was black, 14 point, Times new roman, at 100% scale.

The results of Mangen, Walgermo, and Bronnick’s experiment indicated that reading linear narrative and expository text on a computer screen leads to poorer reading comprehension than reading the same texts on paper. Furthermore, their results suggest that implementing both reading assessment tasks in the same medium - the computer leads to additional cognitive costs.

***P-Books vs. Ebooks: Are There Education Issues?***

In Abram’s article, *P-Books vs. Ebooks: Are There Education Issues?,* he compares printed books and electronic formats. In this study, which was conducted in 2010, he focuses on the learning from these two sources. Abram conducted this study in 2010 and at the time he concluded that this debate will only grow. He did find that overall reading has grown in both printed books and ebooks.

The author found when looking at learning styles, is that “we must acknowledge that some learners need the touch and smell experience to lock down comprehension and retention” (Abram 2010). Text and illustrations are embedding in the right place as the author tries to make the reader feel. In adapting e-reader books on a screen this feel gets lost since the focus is getting the print and illustrations to fit on the size of the screen. The ability to change the font size can help many readers young and old with visual problems, but according to Abram, “we know that the layout can enhance learning by stimulating more than one learning style at a time.” Also, in this article, “ the high-contrast typesetting on good paper continues to rank higher in readability. This may change quickly, but e-paper readers have less contrast, and LCD readers such as the iPad are brighter than necessary for text in order to support video and website display.” This article was written in late 2010 and many of these issues have not been tested and is still up for debate.

**Eye movement analysis of reading from computer displays, eReaders and printed books**

In this article, *Eye movement analysis of reading from computer displays, eReaders*

*and printed books* it conducts a quantitative study to compare eye movement while reading three different types of ebooks and using a book. The three different tools were a desktop PC, and eReader and an iPad and of course a printed book. The method of this study was using video-oculographic technology to record eye movements. In the study of using the computer, a camera was placed under the screen of the computer to record the eye movements and in the other tests (using iPad, eReader and printed book) the subject wore the two cameras. One was used to record the movement of his eyes and the other recording the scene in front of him.

The study showed that the behavior from all three ebooks had a similiar reading behavior that when reading from a printed book. The data showed the number of fixations, their duration, and the direction of the movement. The percentage of these recording showed only a small difference when reading from a computer PC. In all the other eBooks and printed book, the results were about the same.

**A comparison of the influence of electronic books and paper books on reading comprehension, eye fatigue, and perception**

This research assessed the usability of paper books and electronic books to measure comprehension, eye fatigue, and perception. A total of 56 sixth graders participated in the study. This study found that there is a significant “book effect” on quiz scores. Compared to e-books, paper books enabled better reading comprehension. Regarding eye fatigue, students had significantly greater eye fatigue after reading e-books than after reading paper books. Students were satisfied with the e-book, but they preferred paper books. The Korean students studied, who have had a higher level of exposure to technology than those in other countries, did not show positive behavioral intentions toward e-books.

The intent of this paper was not to choose between print and electronic materials. Rather, new media such as e-books add to our choices rather than substituting one form for another. The results of this study did verify structural relationships among eye fatigue, comprehension, and the perception of e-books.

**User attitudes toward dedicated e-book readers for reading**

This paper explored what factors drive users to use dedicated e-book readers for reading. The study proposes a causal model that explores how convenience, compatibility, and media richness affect users' attitudes towards the dedicated e-book readers for reading. The results of this study suggest that convenience, compatibility, and media richness all significantly contribute to dedicated e-book reader acceptance.

For studying the adoption of information technology the Technology Acceptance Model (TAM) was used. A quantitative approach is used in this research. A questionnaire with 23 items comprising six dimensions - perceived usefulness, perceived ease of use, convenience, compatibility, media richness, and intention to use - was constructed. This research used the partial least squares (PLS) technique to perform the data analysis. PLS is a regression-based technique that originates from path analysis.

**Assumptions**

In this study we have assumed that all participating students have some basic familiarity with both personal computers and electronic tablets, either through prior experience in the classroom, or at home. We also assume that these students do not have difficulty or problems reading from a digital screen or have any other vision impairment that would be problematic for this study. Finally, though it may not be an everyday occurrence, students do not believe it is uncommon to use computers and tablets in the classroom and they are aware of the expectations while using them.

**Foreshadowed Problems**

There are several expected problems we may encounter in the study. These include a relatively small sample size that may have trouble generalizing and being applied to a larger audience. Student preferences of paper, tablet or computer screen may play into how well they perform or how seriously they take the assessments. Slight text formatting differences between assessments or additional visuals that we cannot modify on the digital assessments could make one assessment easier to read or focus on. Some students may desire to hurry through the digital tests in the hopes of playing or working on something else on the computer. Finally, some students may feel more comfortable looking straight ahead at a computer screen instead of leaning forward to read and write on paper. Other students may feel more comfortable reading from a sheet of paper and using a pencil. Some students, as shown in our literature review, may prefer the portability of paper.

**Anticipated Results**

The anticipated results from our study are that students will perform better reading from printed text than a digital format. Reading from digital tools creates many new problems never

before measured. Such problems may be LED lighting from screens, contrast, and the overall

feel of not having a book affecting the comprehension of reading materials, overall

learning, and better testing.

**Definition of Terms**

*Digital text:* Any text-based information that is available in a digitally encoded human-readable format and read by electronic means.

*Digital native*: A person born or brought up during the age of digital technology and therefore familiar with computers and the Internet from an early age.

*Quasi-experimental design:*  An experimental design that answers a cause and effect relationship and has an intervention controlled by the experimenters.

*Reading comprehension:* The ability to read text, process it and understand its meaning.

*Smarter Balanced Assessment Consortium:*  Assessments aligned to the Common Core Standards that are *both summative assessments for accountability purposes and optional interim assessments for instructional use.* They use computer adaptive testing technologies that provide actionable data.

*Study Island:* A web based instructional and diagnostic tool that provides state standards practice. It is subscription-based and can be accessed through any device that supports internet connectivity.

**Significance of the Proposed Study**

Until recently, students have been taught and assessed by means of paper and pencil. However, technology now allows for digital assessments that can be graded much more quickly and are capable to adapt so they can more clearly identify a student’s ability level. Due to these and numerous other reasons, California is now using computer based Smarter Balanced testing for the subjects of math and English language arts, students will be expected to demonstrate their reading comprehension by reading from a digital screen and answering on a computer.

While technology is integrated into almost every classroom, teachers currently spend most of their time instructing students without the aid of individual student computers by using books, paper and pencils while expecting students to translate their skills to a new medium at the end of the year. Though teachers are developing new common core methods to teach, these instructional changes may or may not be sufficient to help student achievement on the new digital assessments. If there is an inherent difference between students reading comprehension when reading from a screen instead of paper, additional changes will need to be made to teaching strategies.

**Design and Methodology**

The sample group, or participants in this study, were 50 third grade students. These students were from the Hesperia Unified School District, and the Banning Pass area of California. Of this study, 27 of the participants were girls, and 23 were boys. Ages range from 8 to 10 years of age. Socioeconomic backgrounds varied dramatically. Some students were both socially and economically disadvantaged, some only socially disadvantaged with high incomes, and others had neither disadvantage.

The materials for the study consisted of two different mediums. All students were given two reading comprehension assessments. First, students were given a paper and pencil assessment printed from a subscription website titled, *Study Island*. This assessment was at a third grade level. It consisted of multiple paragraphs and ten question. The format of the assessment was multiple choice. Secondly, all 50 students were given an online reading comprehension assessment from the same website, *Study Island*. This assessment followed a similar format. However, this online assessment was delivered in different formats. Some students took the assessments on Personal Computers, and some took the assessment on tablets. The format of the online assessments was full color, and multiple choice.

The paper and pencil assessments were prepared into packets of five pages. These assessments were black and white. Each student was given an assessment and it was explained that this was not a timed test. Students were told to do their best and read every question. The online assessment required that students log in to Study Island. Once logged in, students needed to select the correct assessment. Once selected, students could begin work on the ten questions. All students were given comparable directions to the paper/pencil assessment.

Two of the researchers in this study taught third grade. In designing this research study we decided it to be most beneficial to measure the progress of students against themselves. Because all student took both assessments we were able to measure the mean and standard deviation, for the group. These data were compared to determine a difference between the online and paper/pencil assessments. Data analysis also included a regression analysis, a coefficient of determination, and a T-Test.

Measurements for this study are reported in the number of correct responses students answered. In short, students scores on both assessments were out of 10 possible points. These scores were converted into percentages. Once the percentages were obtained the mean for each group was calculated. Having the mean score provided the opportunity for further data analysis. This analysis is explained in greater detail later in the paper.

**Instrumentation and Data Collection**

*Test Group1:*

Paper exam for test group 1 - The standard #2 pencils were used to fill in answers on the answer sheet. The information for the paper exam was printed in black ink using standard Arial 12pt single space font. Each test was presented to the students on 20lb thickness, 8.5x11 size, 100 brightness (based on the U.S. Tappi Scale) paper. Each student had a copy of the exam and completed a total of 10 multiple choice questions. No time limit was given, and the students were only allowed to take the test once.

Digital exam for test group 1 - Kindle Nook Tablets were used to access the Study Island Website using a wireless data network connection. Each student was given a segment to read prior to choosing from several multiple choice questions. Each question is timed and gives instant feedback to the student. The students know how well they are performing in the test during the exam. The test was timed, students were expected to finish the test with a reasonable amount of time, however, time was not a determining factor. Once all of the exams had completed, the data was then downloaded in the form of a Microsoft Excel spreadsheet.

*Test Group 2:*

Paper Exam for test group 2 - Paper exam for test group 1 - The standard #2 pencils were used to fill in answers on the answer sheet. The information for the paper exam was printed in black ink using standard Arial 12pt single space font. Each test was presented to the students on 20lb thickness, 8.5x11 size, 100 brightness (based on the U.S. Tappi Scale) paper. Each student had a copy of the exam and completed a total of 10 multiple choice questions.

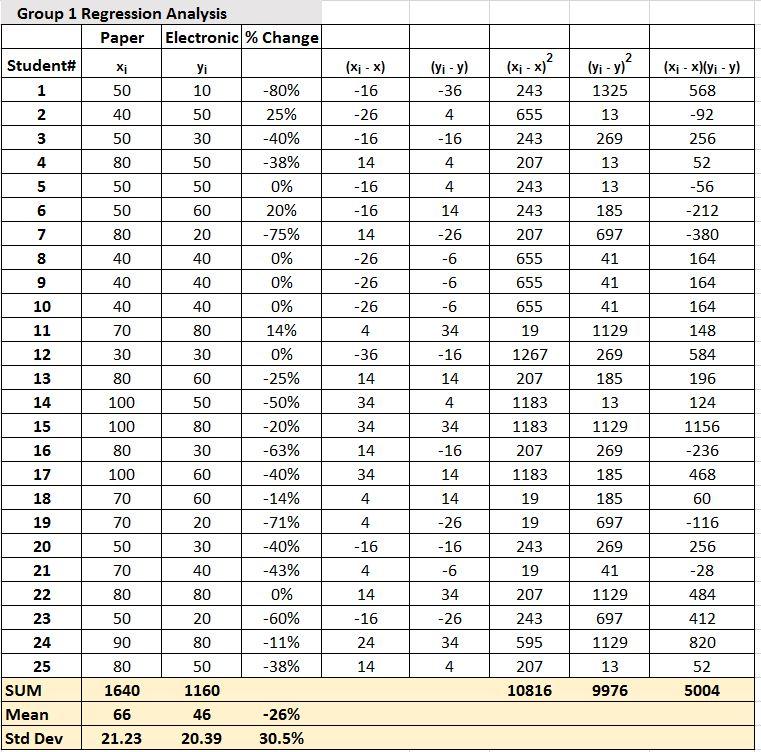
Digital Exam - 3 Apple Computers, standard Apple keyboards, and 17in LCD monitors. The default settings on brightness and contrast were used during the online tests which were conducted on the *Study Island* website. Students were rotated through during this class’s center time due to a lack of personal computers for every student. The tests were timed internally, however, this was not a determining factor in the outcome of the experiment.

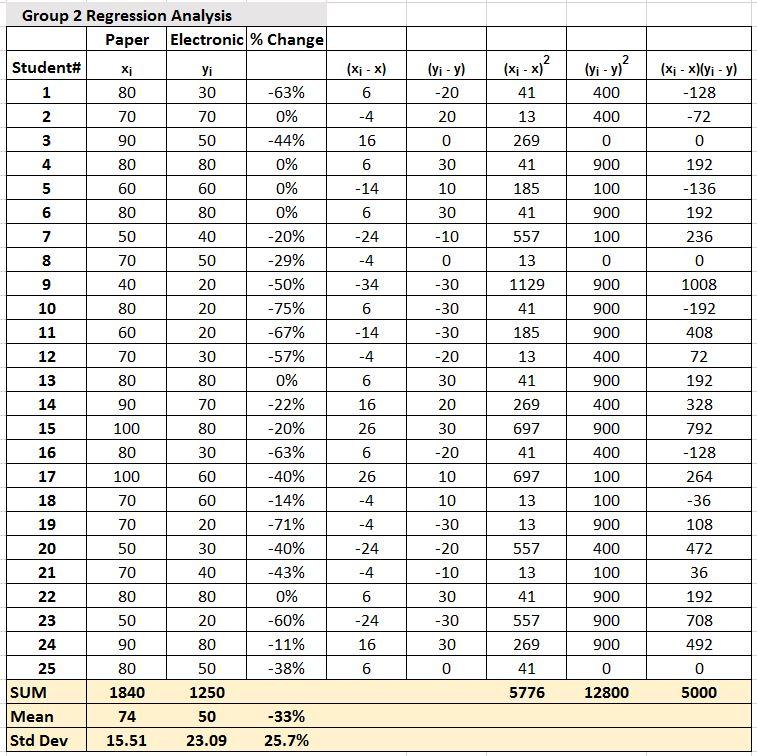
**Data Treatment Procedures**

The data from the digital exams were compiled from raw information associated with the students actual performance by the *Study Island* website, this raw data was then downloaded in the form of a Microsoft Excel spreadsheet, then processed and compared.

The data from the paper exams were recorded and compiled into the same type of data that the digital exam data produced.

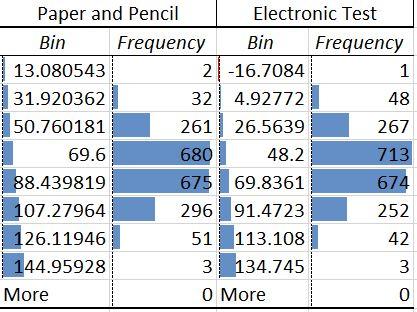
**Presentation of the Findings**





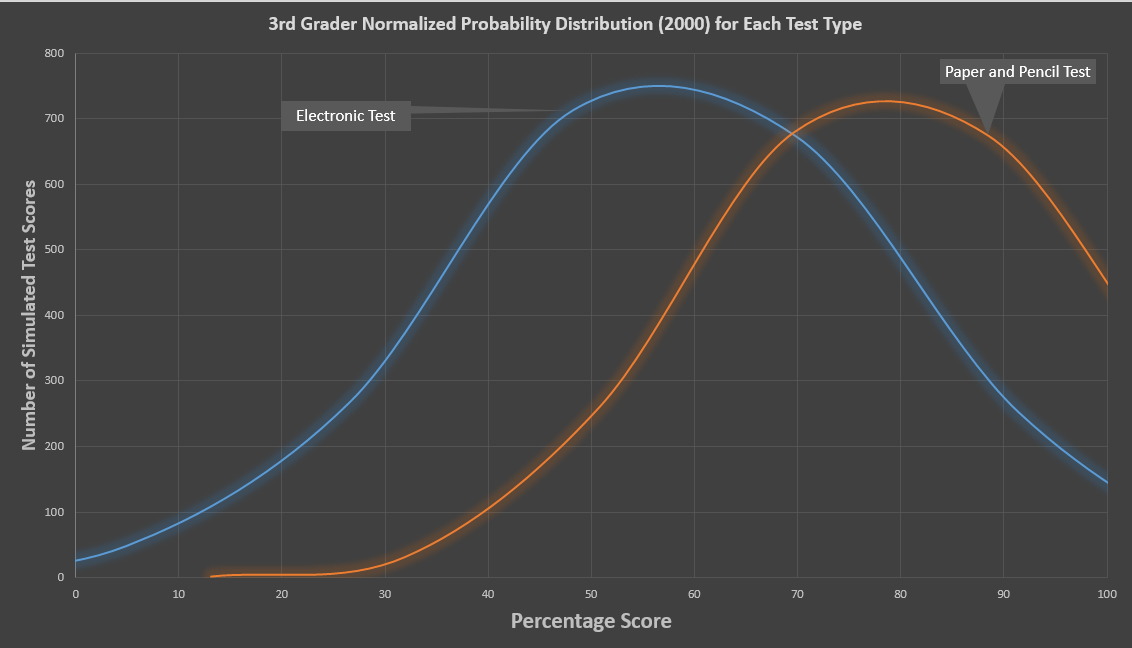
The raw data was then passed through a series of formulas to discover the significance of the impact online testing had versus paper testing. The methods used to derive the percentage of impact required 2 main formulas; the regression equation, and the coefficient of determination. Using the regression formula ŷ = b0 + b1x, first required the summation calculation b1 = Σ [ (xi - x)(yi - y) ] / Σ [ (xi - x)2], and for b0 = y - b1 \* x. With our data these formulas are b1 = 5004/10816 = 0.463 (Group 1), and b1 = 5000/5776 = 0.866 (Group 2) This derived a regression equation of ŷ = 15.442 + 0.463x for Group 1, and ŷ = 14.084 + 0.866x for Group 2.

Once the regression equation was discovered we could then use that information to derive the coefficient of determination (the ratio of the explained variation to the total variation.) The formula used for this requires two standard deviation determinants. Using the standard deviation formula σx = sqrt [ Σ ( xi - x )2 / N ], σy = sqrt [ Σ ( yi - y )2 / N ] and with our data ended up becoming σx = sqrt( 10816/25 ) = sqrt(432.64) = 20.8, σy = sqrt( 9976/25 ) = sqrt(399.04) = 19.98 (Group 1), and σx = sqrt( 5776/25 ) = sqrt(231.04) = 15.2, σy = sqrt( 12800/25 ) = sqrt(512) = 22.627. The final results for the analysis for each group was then formulated as R2 = { ( 1 / N ) \* Σ [ (xi - x) \* (yi - y) ] / (σx \* σy ) }2. The formula was presented as R2 = [ ( 1/25 ) \* 5004 / ( 20.8 \* 19.98 ) ]2 = ( 200.16 / 415.584 )2 = ( 0.482 )2 = 0.232 (Group 1), and R2 = [ ( 1/25 ) \* 5000 / ( 15.2 \* 22.627 ) ]2 = ( 200 / 343.93 )2 = ( 0.581 )2 = 0.338 (Group 2.)

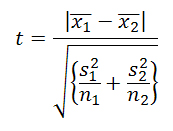
The analysis resulted in two percentages which represents the determinants of a performance degradation that students in this age group with theses circumstances will have on tests taken online versus using paper and pencil. The percentage of correlation for Group 1 is 23.20% and Group 2 is 33.80%. When these two sample groups are averaged, the percentage is 28.50%. Based on these findings we believe that given the same circumstances, the average third grade student will have a 28.5% test performance degradation in online tests versus paper and pencil. 

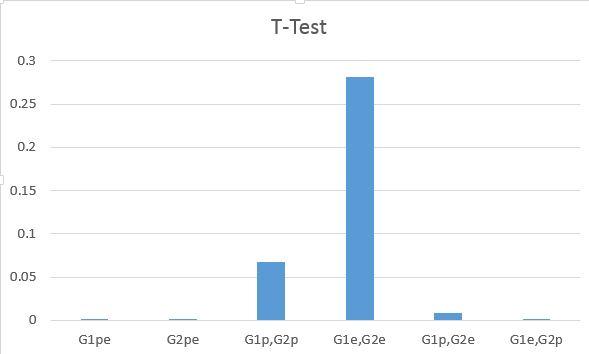
It was then necessary to create a histogram derived from random (computer generated) sample data based on the mean, standard deviation, and a total of 8 deviation distributions (bin) to cover all sample data of each sample group; the following line graph depicts the difference between the combined simulated results of 2000 students (compared against themselves) on paper versus electronic test.

***Normally distributed probability as means and standard deviations is used here is for descriptive purposes.***



A T-Test was also performed, the results are as follows based on the following formula:





Group 1 compared against its own two samples (electronic versus paper and pencil) gave a value of 6.88151E-05, while Group 2 against its own two samples gave a value of 9.41755E-07, Group 1 T-Tested with Group 2 paper and pencil test gave a value of 0.067372134, Group 1 T-Tested with Group 2 electronic test gave a value of 0.280875797, Group 1 paper pencil T-Tested with Group 2 electronic test gave a value of 0.008215003, and the inverse of the previous test Group 2 electronic test T-Tested with Group 2 paper test gave a value of 1.3987E-06. Conclusion of the T-Test is that we can safely reject the null hypothesis (significance level below 5%) for all comparisons except for G1p,G2p and G1e,G2e which proves an overwhelming correlation between the two 3rd grade groups in terms of their performance of each exact test.

**Limitations of the Design**

There are many dynamics which could have contributed to the differences in results. Although the standard in the digital exam was the website and the questions asked, we did not have access to the same visual digital delivery method (screen). The Kindle Nook has a slightly different screen than a standard LED computer screen used in for the group 2’s experiment. Although they are similar, the differences could have played a factor in the outcome of the assessments. Furthermore, the students themselves provide for limitations in the study.

Our sample group totalling 50 students was relatively small, and not representative of large sample group which could have different results. Ideally, all students in this study would have taken the assessments under the exact same conditions. Realistically however, the difference in student backgrounds, previous knowledge, and reading ability, probably contributed to a slight deviation in our findings as compared to a larger sample study. The computer simulation of the normalization data could only use the mean and standard deviation from actual data, not a larger sample group for increased accuracy.

**Conclusion**

As 21st century learning progresses well into the digital age, the way information is organized, processed, and used is changing. Districts, teachers, and students are presently wrestling with the shift from traditional printed text to digital forms of presentational content as well as assessments. Teachers are taught to constantly reflect on their practice and are not left to reflect on the delivery of material through digital forms. Smarter Balanced assessments are delivered through digital forms ands require students not only to comprehend content, but to use a new skill set to use the content as well. The concern for the unknown effects digital devices have on learners, and how effective digital forms are in delivering and assessing content has many questioning if the digital move is the wisest decision. In this study, students were assessed using both digital and traditional forms of text. Looking at the groups used for this study, it can be said that students are 28.5% more probable to perform worse on an assessment that requires students to read and comprehend digital forms of text than on a traditional print text assessment. Should this small sample size of students used in this study be representative of students statewide or even nation wide, the practices now being implemented should be questioned. Are digital assessments not best practices in working with students or will further training and investigation be required before implementing digital assessments for all students?

**Recommendations for Further Research**

Further research will need to be done with an increased sample size and variety of grade levels to verify the reading comprehension results. This study was composed of a relatively small sample size of only 50 students in mixed socioeconomic classrooms at the third grade level. Additional research will help to determine if the results hold true amongst a wider sample size, other grade levels, and students of varying socioeconomic levels.

Research may be most beneficial to focus on digital reading on personal computers or laptops similar to those found in administering the Smarter Balanced assessments, which includes the ability to modify text size, highlight, create notes and other Smarter Balanced test features. Comparing results of paper and pencil assessments versus a Smarter Balanced style digital assessment relates most directly to the classroom activities students are currently seeing versus the state’s year end assessment.

**Work Cited Page**

Abram, S. (2010). P-Books vs. Ebooks: Are There Education Issues?. *Multimedia &*

*Internet@Schools*, *17*(6), 13-16.

Aydemir, Z. & Ozturk, E. (2012). The effects of reading from the screen on the reading motivation levels of elementary 5th graders. *The Turkish Online Journal of Educational Technology, Volume 11*, (Issue 3). Pp. 357 - 365

Hanho, J. (2012) "A comparison of the influence of electronic books and paper books on reading comprehension, eye fatigue, and perception", The Electronic Library, Volume 30 Iss: 3, pp.390 – 408

Jung-Yu, L., & Chih-Yen, C. (2011). User attitudes toward dedicated e-book readers for reading. *Online Information Review, 35*(4), 558-580.

Larson, L. C. (2013). It's Time to Turn the Digital Page: Preservice Teachers Explore E-Book Reading. *Journal Of Adolescent & Adult Literacy*, *56*(4), 280-290.

Lauterman, T., & Ackerman, R., (2014). Overcoming screen inferiority in learning and calibration. *Computers in Human Behavior, Volume* 35, 455-463

Mangen, A., Walgermo, B. R., & Bronnick, K. (2013). Reading Linear Texts on Paper versus Computer Screen: Effects on Reading Comprehension. *International Journal Of Educational Research, 5861-68*

Spencer, C. (2006). Research on learners’ preferences for reading from a printed text or from a computer screen. *Journal of Distance Education, Volume* 21 (Issue 1), pp. 33-50.

Sun, S. , Shieh, C. , & Huang, K. (2013). A research on comprehension differences between

print and screen reading. *South African Journal of Economic and Management Sciences*, *16*(5), 87.

Zambarbieri, D., & Carniglia, E. (2012). Eye movement analysis of reading from computer

displays, eReaders and printed books. *Ophthalmic & Physiological Optics*, *32*(5),

390-396.