St. John's University

St. John's Scholar

Theses and Dissertations

2020

MOVING TOWARDS COMPUTER ADAPTIVE TESTING: THE EFFECT OF EXPERIENCE WITH TECHNOLOGY ON ELEMENTARY STUDENTS' SCORES AND ATTITUDES

Brittany Neligan

MOVING TOWARDS COMPUTER ADAPTIVE TESTING: THE EFFECT OF EXPERIENCE WITH TECHNOLOGY ON ELEMENTARY STUDENTS' SCORES AND ATTITUDES

A dissertation submitted in partial fulfillment
of the requirements for the degree of
DOCTOR OF EDUCATION
to the faculty of the Department of
ADMINISTRATIVE AND INSTRUCTIONAL LEADERSHIP
of
THE SCHOOL OF EDUCATION
at
ST. JOHN'S UNIVERSITY
New York
by
Brittany A. Neligan

Submitted Date March 10, 2020 Approved Date March 10, 2020

____________________________ _______________________________

Brittany A. Neligan Dr. Rene S. Parmar

ABSTRACT

MOVING TOWARDS COMPUTER ADAPTIVE TESTING: THE EFFECT OF EXPERIENCE WITH TECHNOLOGY ON ELEMENTARY STUDENTS' SCORES AND ATTITUDES

Brittany A. Neligan

In this quantitative study, students' growth over the course of the school year on the i-Ready test were analyzed. Using an ex post facto design, the i-Ready growth scores of students with experience of the testing format (n=45) were compared to the growth scores of the students with no experience of the testing format (n=179). A descriptive analysis was performed to analyze the students' feelings and perceptions about adaptive Computer-Based testing conducted within their schools. Fourth and fifth grade students (n=27) answered an open-ended survey, which were used to see how elementary school students feel about the shift from Paper-Based to Computer-Based testing. Results indicate that there were no significant differences in scores between students with experience and students without experience, nor were there differences between the achievement of students based on gender or instructional groups. The surveys indicate that students enjoy using computer-based testing, but experienced trouble with navigating through the tests, efficiently using tools, and implementing other self-regulatory behaviors that they often use when working on paper-based tests. This study indicates that more instructional time needs to be spent using computers, in order to teach students self-regulatory strategies that can help students to become more comfortable and adept with computer-based tests. With more explicit instruction, student growth on various assessments may increase.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I:

INTRODUCTION

With educators encouraging and developing the 21st Century Learner, it has become apparent that computer utilization and technology cannot be ignored in the classroom. Instruction has become infused with technology, including Smart Boards, Google Classrooms, 1:1 devices, and much more. The utilization of technology in the classroom has led educators to explore and purchase Computer-Based Testing (CBT), which is an alternative assessment instrument used to supplement the traditional paper-based tests, or PBTs (Jeong, 2014). Computer based tests include Computer Adaptive Testing (CAT), which is a unique form of assessment that adapts to a student's ability level (Shapiro & Gebhardt, 2012), as well as benchmark and summative assessments. High-stakes testing is moving toward a computer-based format, as seen in the initiatives by many state boards of education, including New York State, the site of the present study.

The first section of this literature review includes information about Self-Regulated Learning, and how this theory relates to Computer-Based Testing. Following that is a brief review of research on the use of Computer-Based Testing, including both supportive and contradictory studies.

Rationale of Study

The present study extends the existing state of knowledge to include research that examines the way by which teachers are preparing their students and utilizing technology in the classroom to best prepare their students for Computer-Based assessments. The study analyzed how students feel about Computer Adaptive Testing that is given in their school throughout the academic year.

Schools are beginning to embrace 21st century learning skills and implement technology, and they are doing so at a fast rate. Among the changes within the school system brought about by advances in technology is the increasing adoption of computer-based assessment for diagnostic purposes, summative evaluation, and high stakes decision-making. More needs to be investigated regarding the effect of computer-based testing, especially when one considers the impact that test-taking has on students and their trajectories into college and beyond. Computer-based assessments have been used at an increasing rate for many reasons, including their ability to assess students immediately, which provides teachers and students with immediate feedback. They also improve test administration, decrease testing expenses, and reduce paper consumption (Chua, 2012, p. 1580; Jeong, 2014, p. 410).

With the push for computer-based assessments, many states are beginning to adopt state wide, standardized computer-based assessments. In fact, the shift has begun in New York, the site of the present research. After piloting the CBT state assessment in 2016 and offering the option for Computer-Based testing between 2017 through 2018, the state decided, “The goal of the Department is that all Grades 3-8 testing will be delivered on computers by 2020” (New York State Education Departments, 2019, www.nysed.gov).

Unfortunately, the transition from paper-based testing to computer-based testing has not been smooth. According to Brody (2018), the testing company Questar has been given a five-year $44 M contract with New York State to develop the computer based assessments. However, in April, 2018, there was a technical problem in which, “students in certain grades at 263 districts experienced delays, and more than 49,900 pupils completed computer-based tests [later that week].” This left many students and schools stressed, and it left teachers and parents questioning the transition.

Purpose

With state-wide computer-based testing, it is important for research to be conducted at the elementary level to understand the process of implementation as well as the ability of students to perform successfully. Little research has been done thus far, and the models of assessment proposed tend to be based on numerous assumptions about format, ease of administration and use, and congruence with paper-pencil testing, without direct empirical evaluation. The purpose of the study is to:

1. Analyze the growth scores of 3rd and 4th grade students by comparing their growth over the course of the first year using the English Language Arts *i-Ready* computer adaptive diagnostic assessment to their scores in the second year of i-*Ready* assessment implementation.
2. Analyze the growth scores of students who have had one year of experience with the English Language Arts *i-Ready* computer adaptive diagnostic assessment with their peers who are taking the test for the first time.
3. Analyze the effect of gender and instructional program on the growth scores of students taking the English Language Arts *i-Ready* computer adaptive diagnostic assessment.

4. Analyze the students' perception, motivation, and feelings about Computer Based Test after the students complete a practice version of the state exam, provided by the New York State Education Department.

This goal of this study was to help educators to better understand if experience and exposure with testing formats impacts student growth and examine if students are prepared for the state-mandated shift in assessments.

The Shift Toward Computer-Based Assessment

As with all new shifts, increased instruction (specifically strategy-based instruction) is needed to prepare students for the challenges faced when encountering a new test format. For example, one major difference between computer-based and paper-based testing is that on a paper-based test, students have the entire test in their hands throughout the test duration, and they can mark-up the questions, underline, or eliminate choices. Computer-based tests, depending on the testing format, may not offer such functionalities (Boevé, Meijer, Albers, Beetsma, & Bosker, 2015, p. 3). Readability of the digital text is a concern of educators, including students' ability to generalize across instructional materials. Additionally, students have less opportunity to interact with the text, including highlighting and annotating (Worrell, Duffy, Brady, Dukes, & Gonzalez-DeHass, 2016, p. 267). Therefore, students should be exposed to computer-based test practices and various formats, so that they can develop ways to overcome some of the challenges of new testing format. For example, some computer-based tests have a "flagging" option. Navigating the test options before a test may be helpful for many students.

Other skills and abilities are important besides academic strategies. Computer literacy is important for students to navigate a computer-based assessment and should be evaluated before students initiate a computer-based exam. The International Society of Technology Education (ISTE) has developed computer technology and literacy standards for students. Of the seven standards, three would be needed for students in order to complete a computer-based assessment:

1. Become an Empowered Learner, who can…
    - set personal learning goals, develop strategies leveraging technology to achieve them and reflect on the learning process itself to improve learning outcomes.
    - use technology to seek feedback that informs and improves their practice and to demonstrate their learning in a variety of ways.
    - understand the fundamental concepts of technology operations, demonstrate the ability to choose, use and troubleshoot current technologies and are able to transfer their knowledge to explore emerging technologies.
2. Become a Knowledge Constructor, who can…
    - evaluate the accuracy, perspective, credibility and relevance of information, media, data or other resources
    - curate information from digital resources using a variety of tools and methods to create collections of artifacts that demonstrate meaningful connections or conclusions
    - build knowledge by actively exploring real-world issues and problems, developing ideas and theories and pursuing answers and solutions.

3. Become a Computational Thinker, who can…
    - collect data or identify relevant data sets, use digital tools to analyze them, and represent data in various ways to facilitate problem-solving and decision-making.
    - break problems into component parts, extract key information, and develop descriptive models to understand complex systems or facilitate problem-solving.

    (International Society of Technology Education [ISTE], 2019)

Schools and teachers need to be deliberate in teaching students computer-literacy skills in the early primary grades, so that they are ready to use the computer functions, identify problems, extract data, evaluate problems and solve. "[Clearly,] fluency with computer technology goes beyond traditional notions of computer literacy. Computer technology literacy enables one to accomplish a variety of different tasks and in different ways" (Chang, 2008, p. 623).

Significance of the Study

More needs to be investigated regarding the effect of computer-based testing, especially when one considers the importance of computers in our everyday life. Students in states such as Rhode Island and Illinois, as well as in Baltimore County, Maryland, are being given high-stakes standardized tests online (ie, the PARCC English Language Arts Exam). In fact Rhode Island's results for the PARCC exam in its first year of implementation found that "42.5 percent of the students who took the PARCC English/language arts exam on paper scored proficient, compared with 34 percent of those who took the test by computer…[which could be] due in large measure to varying degrees of 'student and system readiness for technology'" (Herald, 2016, p. 1). Nationwide there is a movement toward increasing the administration of high-stakes tests via computers or tablet devices.

Some schools, especially elementary schools, do not have as much access to 1:1 devices as those in high school. Students' age, experience, maturity, and ability to self-regulate may compromise their scores and perceptions during testing. Additionally, teachers' feelings and attitudes may impact the effectiveness of the assessments.

The present study adds to the literature and dialogue on computer use for high stakes assessments by discussing differences in summative assessments and adaptive testing that educators must understand if they are to make useful interpretations of the data. Connections with theories of student learning, motivation, and self-regulation are incorporated into the discussion. The study provides insight into issues surrounding test administration that can be of use to educators and administrators who are considering wide-spread implementation in their schools. From the students' perspective, the study reveals usage of test-support tools by test-takers and provides test design considerations. Finally, the study contributes to policy discussion on acceptance of and implementation of computer-based assessments.

Definition of Terms

Curriculum-Based Measure (CBM) act as a summative or ongoing assessment. Scores obtained by students on Curriculum Based Measure identify student performance or concept development in comparison to grade level expectations (Shapiro & Gebhardt, 2012).

Computer-Adaptive Test (CAT) are tests that refine the selection of items based on a student's response and help teachers by diagnosing students' areas or strength and weaknesses (Shapiro & Gebhardt, 2012).

Summative Assessment *"*uses data to assess about how much a student knows or has retained at the completion of a learning sequence" (Dixson & Worrell, 2016, p. 153).

High-Stakes Testing is a name used to describe norm-referenced tests that are used to compare one's individual score to a large group of test-takers. Such test are usually given nationally or state-wide and are often used to evaluate students, teachers, schools, districts, ad states. High stakes tests often have universal test administration and directions, as well as a set amount of time for each test taker (Merchant, 2004, pp. 2, 3).

Conclusion

The goal of this study is to examine the impact of experience with Computer Based Testing, as it may indicate that exposure to the computer-based testing format, such as the *i-Ready* program, may lead to increased performance. In addition, the goal of this study is to examine the perceptions of students that influence Computer-Based Testing at elementary level. If there are negative feelings towards the CAT, it should encourage educators and administrators to reflect and ask if computer-based testing is right for students of all ages, or if it is better-suited for students of a certain age.

CHAPTER II:

REVIEW OF LITERATURE

This chapter presents a short summary of the research on computer-based assessment, particularly as it relates to summative or high-stakes tests. To begin, theoretical perspectives that undergird the assumptions of computer-based testing is reviewed. Next, a look at prior studies that examine and compare CBT approaches is presented, followed by research on student experiences and perceptions. A report of studies that have raised questions about the implementation and interpretation of CBT is included. The chapter concludes with a statement of how the present research builds upon prior studies and extends the research-base on CBT.

Theoretical Framework

One aspect of this research examines the effect of experience with Computer Based Tests and how it may impact student growth. Bruner's Theory of Constructivism includes student readiness and scaffolding. Information must be introduced to students at an appropriate age and developed over time. Therefore, Bruner felt that teachers should use a spiral curriculum, in which students are introduced to content and skills and then revisit content to better develop their understanding (Schunk, 2016, p. 310). Vygotsky's Zone of Proximal Development (Schunk, 2016, p. 314) expands on this concept, whereby students can learn new content but may need guidance from adults or peers to accomplish a task. "The experiences one brings to a learning situation can greatly influence the outcome" (Schunk, 2016, p. 315). These theories indicate that students may need practice with and guidance from teachers and peers before taking Computer-Based Tests.

Self-regulated learning is a vital element of student development. This means, being involved in one's learning and performance on a multi-dimensional level, including behaviorally, cognitively, metacognitively, and motivationally (Schunk, 2016, p. 398). Self-regulated learning (SRL) is multi-faceted and includes self-monitoring and self-reinforcement.

The model of Self-Regulated Learning developed by Zimmerman and Moylan in 2009, the "Cyclical Phases Model" (Panadero, 2017) illustrates the thinking that is needed to complete adaptive tests and to grow over the course of the academic year (Figure 1).

According to Zimmerman and Moylan (2009), the model in Figure 1 depicts that self-regulation includes not only strategy and time management, but also self-consequences and metacognitive monitoring. After the performance, students should exhibit self-judgement and self-assessment, which should lead to forethought for future performances. This can include goal setting and planning for future assessments (Panadero, 2017). In many curriculum-based tests, the forethought process may be less valuable because tests on the same topic (or chapter, in elementary schools) are not going to take place, as the teachers most often move on to a new chapter and do not test old materials. However, with computer-adaptive tests, this forethought and goal-setting can be very important to the students' growth, as the content may change but the strategies used by the students might improve over time and contribute to their growth.

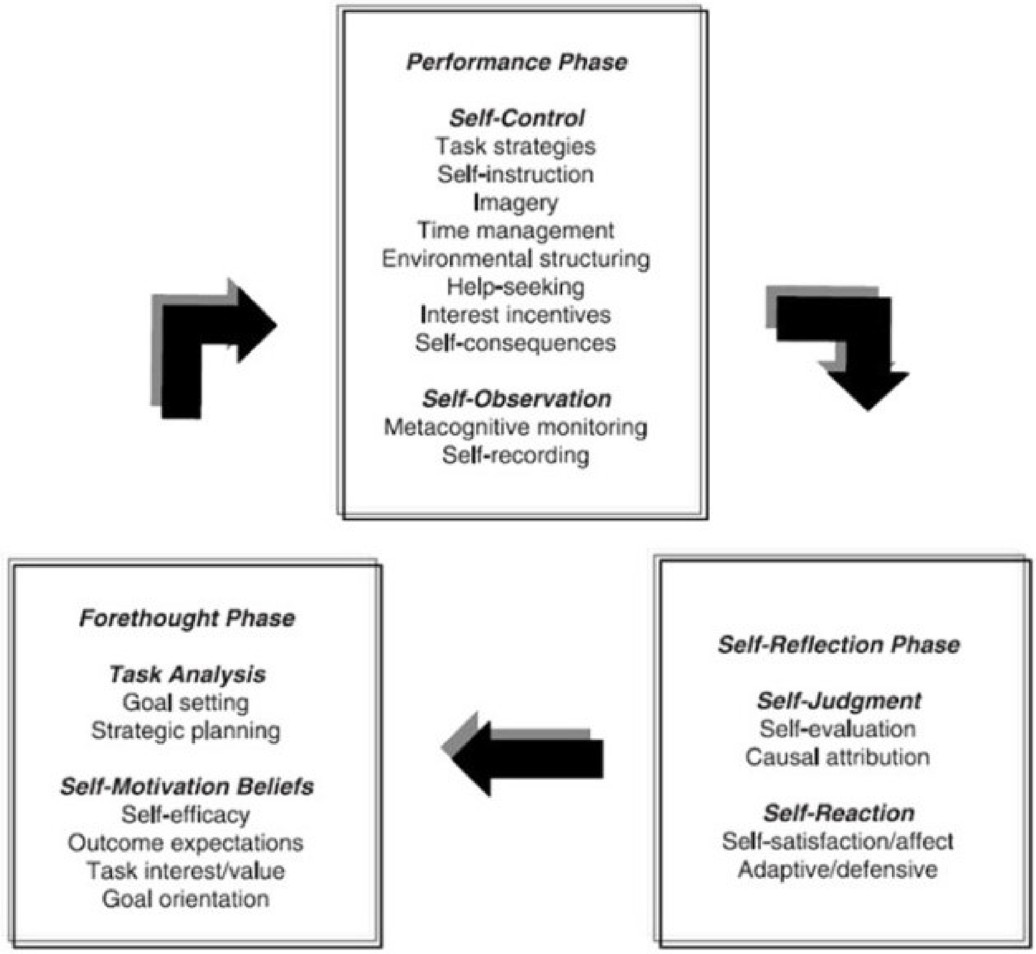


Figure 1. Current version cyclical phases model. Adapted from Zimmerman and Moylan (2009, as cited in Panadero, 2017).

When using the Computer Adaptive Test (CAT), students may find that self-regulation is easier to maintain because of the adjustment of the questions based on their ability. The adjustment of difficulty, on the other hand, may cause student frustration as students are given more rigorous questions, which may also encumber performance. Likewise, Computer Based Tests that are summative, such as the state tests or unit tests, self-regulation may be more challenging for students who are struggling, as the questions do not adjust to meet the capabilities of the students. According to Greene, Moos, and

Azevedo (2011), “Students who are effective at self-regulating their learning will continue to capitalize on the opportunities of computer-based learning environments (CBLE), while those who lack this ability will find themselves at a serious disadvantage. Educators would do well to consider preliminary and formative assessments of their students’ SRL skills, knowledge, and motivation while using CBLEs and then design scaffolding interventions accordingly” (p. 113). Without self-regulated learning skills, students’ achievement on assessments may be hindered, especially when using a new format of testing, such as CBT.

The purpose of this study is to compare growth scores of students and the perspectives of students whose school has started to use adaptive test, *i-Ready*. Therefore, one must consider the constructivist theory, by which people develop their knowledge and understanding through interactions with persons and situations. Constructivism also proposes that one’s learning is influenced by one’s own environment (Schunk, 2016, p. 298). When considering the implications of constructivism, it is important for educators to allow students to interact with computer tools and computer-based assessments in order to develop a deep understanding of the expectations and format. Without the experience of computer-based assessments, student achievement may be hindered. With the shift in assessments, it is important to see how the new trend and experience with a program impacts growth scores, attitudes, and motivation of those taking the tests.

Studies on Computer-Based Testing

There are many ways to incorporate technology into the classroom. Many of these modern technological utilizations, including one-to-one devices, help to promote student success. When teaches embrace the technology, learning can flourish. In a case study conducted by Grant et al. (2015), nine K-12 teachers from various states taught using mobile computing devices or had students in class who used mobile computing devices (MCDs). Researchers then conducted interviews with the participants to find out the teachers' perceptions and feelings of the technology integration. It was found that the use of MCDs enhanced the classroom experience in many ways. Many teachers used MCDs to supplement their curriculum. They incorporated aspects of Project Based Learning, including, [using] device applications, communicating with others, recording video and audio, projecting and displaying work, and creating news casts (Grant et al., 2015, p. 41). The research regarding the success of computer-based instruction and mobile computing devices may help to persuade administrators, educators, and policy-makers into using computer-based assessments more regularly.

According to Pittman and Gaines (2015), "As students begin to develop technology habits, it is vital to teach them how to effectively use the tools available to them in a safe and ethical way, and this is only possible when there is a robust level of technology integration in classroom instruction" (p. 542). For this reason, using devices in classrooms has grown in popularity, but it is important to note the differences between integrating technology into the classroom and using technology as an assessment tool for high-stakes tests. Students who are in schools with devices should access computers for information, communicate, and practice academic skills in order to reinforce what is taught by a teacher, *as well as* practice using assessment technology (Pittman & Gaines, 2015).

A study by Zhang, Trussell, Gallegos, and Asam (2015) it was indicated that when students in a fourth-grade classroom used math apps, including SplashMath, all student improved. More notable, however, is that the achievement gap closed between the struggling students and their higher-achieving classmates. Using these apps also increased student engagement and student practice. Students were given immediate feedback and tracked their progress according to their results (Zhang et al., 2015, p. 38). Clearly, there are benefits to using such devices and with the growing use of computer-based assessment, it may be helpful to begin using more technological tools, such as SplashMath, within the classroom to help students adapt to the new expectations.

However, the programs used to instruct students in basic skills do not mirror tests like the *i-Ready* adaptive test. When students take adaptive, computer-based tests, they are sitting for longer periods of time (up to two hours), immediate feedback is not given by the high-stakes tests, and the students are not always working (or practicing) skills from their curricula. Questions can reflect skills that are cumulative, from previous grades, or may be accelerated as students progress through. The differences between the project-based learning that exists in classes with devices and the testing that is beginning in schools is significant, in that student experiences with computers does not dictate that students are be ready for computer-based assessments. Pittman and Gaines (2015) suggest the importance of showing students how to use computer ethically and effectively (p. 542). Therefore, teachers and schools may need to take more time to show students how to effectively use and take the various types of assessments that are now being used on the computers, including standardized and adaptive tests.

For some students, practice and exposure to the test format may be enough exposure for students to be ready for the assessment shift, but for struggling readers and/or students with disabilities, more direct instruction is needed. For example, in a study conducted by Worrell et al. (2016), four students were systematically taught the "NRUN" strategy, meaning Number the paragraphs; Read each paragraph; Understand what you read; and Note key words (p. 268). The purpose of the study was to see if the reading strategy NRUN would be used by students when interacting with the text on a computer. With explicit instruction of the reading strategy, the students' computer-based test scores increased. Therefore, students at the elementary level may need explicit instruction from teachers in order to generalize skills that were once performed on paper to skills that are now performed on the computer.

A Comparison of Computer Based Assessment Approaches

Assessments in elementary schools can vary in format. Computer-adaptive testing (CAT) has emerged as a viable option for universal screening. These tests refines the selection of items based on a student's response and help teachers by diagnosing students' areas or strength and weaknesses (Shapiro & Gebhardt, 2012). Examples of this kind of test include the NWEA and STAR assessment. CATs are a formative way of collecting data and help teachers to adjust their instructional decisions based on the data they receive.

Curriculum Based Measures (CBM), on the other hand, act as a summative or ongoing assessment. Scores obtained by students on Curriculum Based Measure identify student performance or concept development in comparison to grade level expectations (Shapiro & Gebhardt, 2012). CBM assessments come in a wide variety and can include unit tests, state test assessments, and much more. CBM have been traditionally given using the paper-based format and are often associated with progress monitoring in schools. Some CBMs are now being conducted on the computer. Table 1 indicates the traits of each kind of assessment for comparison.

Shapiro and Gebhardt (2012) compared the results of CAT and CBM assessments by analyzing the scores of 352 students in grades 1-4 from two different schools in rural Pennsylvania. Indicators of student success in math includes the PSSA, the Pennsylvania state assessment (CBM assessment), the STAR assessment (a CAT assessment), and AIMSweb (Math Concept/Application assessment, MCAP). The results indicate that the three different kinds of test show little correlation due to the variety of domains offered within each test. However, they did reveal that the STAR assessment (CAT) was the best predictor of student scores on future state scores. Furthermore, the results showed that there were distinct differences in data collected through CAT and CBM assessments. This makes it clear that assessments chosen by a school should be well-connected to the core instructional curriculum and should help to organize students into instructional groups easily (Shapiro & Gebhardt, 2012).

Table 1

Kinds of Assessments

Formative Learning Assessment	Formative Diagnostic Assessment	Benchmark/ Interim Assessment	Summative Assessment
What is it?			
Formative learning is the process of teaching students how to set goals for their learning, to identify their growth towards those goals, to evaluate the quality of their work, and to identify strategies to improve.	Formative diagnostic assessment is a process of questioning, testing, or demonstration used to identify how a student is learning, where his strengths and weaknesses lie, and potential strategies to improve that learning. It focuses on individual growth.	Benchmark or interim assessment is a comparison of student understanding or performance against a set of uniform standards within the same school year. It may contain hybrid elements of formative and summative assessments, or a summative test of a smaller section of content, like a unit or semester.	Summative assessment is a comparison of the performance of a student or group of students against a set of uniform standards.
Who is being measured?			
Individual students are measuring themselves against their learning goals, prior work, other students' work, and/or an objective standard or rubric.	Individual students. The way they answer gives insight into their learning process and how to support it.	Individual students or classes.	The educational environment: Teachers, curricula, education systems, programs, etc.
How often?			
Ongoing: It may be used to manage a particular long-term project, or be included in everyday lessons. Feedback is immediate or very rapid.	Ongoing: Often as part of a cycle of instruction and feedback over time. Results are immediate or very rapid.	Intermittent: Often at the end of a quarter or semester, or a midpoint of a curricular unit. Results are generally received in enough time to affect instruction in the same school year.	Point in time: Often at the end of a curricular unit or course, or annually at the same time each school year.
For what purpose?			
To help students identify and internalize their learning goals, reflect on their own understanding and evaluate the quality of their work in relation to their own or objective goals, and identify strategies to improve their work and understanding.	To diagnose problems in students' understanding or gaps in skills, and to help teachers decide next steps in instruction.	To help educators or administrators track students' academic trajectory toward long-term goals. Depending on the timing of assessment feedback, this may be used more to inform instruction or to evaluate the quality of the learning environment.	To give an overall description of students' status and evaluate the effectiveness of the educational environment. Large-scale summative assessment is designed to be brief and uniform, so there is often limited information to diagnose specific problems for students.
What strategies are used?			
Self-evaluation and metacognition, analyzing work of varying qualities, developing one's own rubric or learning progressions, writing laboratory or other reflective journals, peer review, etc.	Rubrics and written or oral test questions, and observation protocols designed to identify specific problem areas or misconceptions in learning the concept or performing the skill.	Often a condensed form of an annual summative assessment, e.g. a shorter term paper or test. It may be developed by the teacher or school, bought commercially, or be part of a larger state assessment system.	Summative assessments are standardized to make comparisons among students, classes, or schools. This could a single pool of test questions or a common rubric for judging a project.

Note. Adapted from Sparks (2015).

Jeong (2014) compared the results of two testing formats in an elementary school in Korea. Seventy-three sixth grade students (38 male; 35 female) were given an 80-question multiple choice test, including Korean language, math, social studies, and science. All questions were presented in the same way on the computer and on paper. The participants took both versions of the test, and the results were compared. Jeong's (2014) research indicates that all participants performed better on the Paper Based test (which was given first). It was also found that there were significant differences is CBT and PBT scores in two subject areas: Korean and science. For males, there was less of a difference in scores between the two testing formats (a slight difference in Korean). The female students, on the other hand, had significantly different scores in all three subject areas: math, science, and Korean (Jeong, 2014, pp. 415-416). These findings indicate that the experience of taking a CBT may be different for boys and girls at the elementary age.

Research on Computer-Based testing has been focused on students at the middle-school, high school, and university level. Results may differ from students at the elementary school level, but Chua's (2012) study was used to help guide the researcher who conducted this current study. Chua (2012) compared Paper-and-Pencil Testing (PPT) to Computer-Based Testing (CBT) at a university level. One hundred forty participants (68 males; 72 females) enrolled in a Malaysian teacher education program were randomly assigned to one of four groups: two treatment groups and two control groups. The treatment groups were given Computer Based pre-tests and post-test; The control group were administered the same tests in the paper-based versions. The results show that CBT was a more reliable measure, reduced time spent taking a test, and increased self-efficacy. This research might encourage schools to begin adopting one-to-one devices and/or computer-based assessments in order to increase in student motivation, the increase in self-efficacy and the reduced amount of test-taking time would be an advantage for *all* students.

In each study, it was shown that there were differences between the students' performances on each version of assessments mentioned. Over time, researchers may find that one assessment format outweighs another.

Student Experience with Computer-Based Tests

Backes and Cowan (2019) conducted a study to find the test mode effect of student familiarity and school administration of tests across the state of Massachusetts, as the state rolled out the PARCC exam on the computer. The study took place state-wide and across three years. The results of this study indicate that there was little mode effect relating to school testing administration in the area of math. Rather most improvements in math scores were related to student familiarity and experience with the computer-based test. In English Language Arts, testing administration did account for a portion of the mode effect, as did student experience. Despite experience impacting student performance, the results still conclude that students who took the paper-based test still performed better than those who took the computer-based test (Backes & Cowan, 2019, pp. 11, 12).

Student Perceptions of Computer-Based Assessments

Richardson et al. (2002) interviewed 24 students who took the World Class Tests, which is an internationally administered exam, and includes computer-based and paper-based portions, assesses math and problem-solving skills, and identifies achievement of gifted and talented students. Of the 24 participants, 21 of them indicated that they preferred using the computer-based portions of the test. Students also preferred the colors and images on the computer, the ease of typing (as opposed to an aching hand after writing on a paper-based test), and the tasks on the computer, which students said were more interesting than the paper-based tasks (Richardson et al., 2002, p. 642). Students' feelings, perceptions, and preferences for computer-based tests may increase motivation during testing and impact future student achievement.

A case study by Özden, Ertürk, and Sanli (2004) surveyed and interviewed 46 college-aged students in the Department of Computer Education (p. 80). Of the students, only four considered their computer experience *poor* (Özden et al., 2004, p. 81). Based on the results, 58% of students liked the immediate feedback; 79% liked the testing format better than paper and pencil; and 92% thought the computer assessments were faster than paper-based tests (Özden et al., 2004, p. 86). Many students agreed that the tools needed improvements, such as note-taking sections or opportunities to revise answers (Özden et al., 2004, p. 88).

Özden et al. (2004) concluded that the key to student perception of online assessments is experience (p. 90), which supports the theoretical framework that practice exposure plays an important role in student success. Additionally, higher-achieving students develop test-taking strategies for the computer assessment faster than their peers who are less academically successful. However, despite training, anxiety about the new test did exist, making a strong point that experience with online tests coupled with a warm environment are *both* key components to more positive student perceptions of online testing. This makes it clear that it is important for educators to be aware of students' test taking perceptions as they roll out and mandate new test-taking formats.

Contradictory Studies on Computer Based Assessment

Some research has been conducted that has not proven to show significant correlations between the format of testing and the success of a student. For example, in a study by Boevé et al. (2015), 401 college-aged participants were randomly assigned to CBT and PBT midterms. Then, they were given the other format for their final exam. After the semester, the students were given a survey on their acceptance of the computer-based version and paper-based versions of the test. It was found that there was no significant difference in the average number of questions answered correctly between the computer-based and paper-based mode for both the midterm and final exam at the post-secondary level. However, the surveys indicated that students felt more positive about their ability to work when working on the paper-based version of the test.

In addition, Jarodzka, Janssen, Kirschner, and Erkens (2015) studied attention splitting when conducting computer-based assessments. Twenty-two pre-university students (1 male; 21 females) in the Netherlands were given the *Art Appreciation* national exam for Dutch secondary education. All tests were computer-based, and researchers analyzed the difference between an integrated test format (wherein all relevant information is on one screen) and a split format (wherein the information needs to be accessed). Results indicated that students performed more efficiently on test items presented in a split format than on items presented in an integrated format.

If there is no significant difference between testing formats, this could allow for more student choice in terms of testing format.

Conclusion

From reviewing the literature, it is clear that there are many different factors to consider when evaluating the success of a testing format, including student success, student perceptions, and experience. Jeong's (2014) study indicated that there were significant differences between the results of paper-based and computer-based tests, whereas Boevé et al. (2015) found that there were no significant differences between the two testing formats. Both Özden et al. (2004) and Backes and Cowan (2019) concluded that experience is integral into the success of computer-based testing and an important consideration in rolling out assessment programs.

Despite the research provided, it is apparent that there are gaps that need to be filled in the area of computer-based testing. The research that has been conducted thus far has focused on secondary and post-secondary students. Little research has been conducted at the elementary level. Very little research has been conducted in the area of supports provided in schools for students, formatting issues, and self-regulatory behavior of students taking the computer-based assessments. While Backes and Cowan (2019) explored test effects, there are few other studies that explore how experience and student background influence or impact student success on computer-based assessments. Lastly, there are no studies mentioned in this literature that include the use of computer-adaptive testing, which are growing in popularity in schools throughout the country. Many studies have focused on summative assessments, rather than student growth and improvement.

Therefore, this study expanded the research that has already been conducted regarding computer-based assessments by focusing on students in younger grades taking a computer-adaptive test, the *i-Ready* test. It also expanded research by Backes and

Cowan (2019) and examine how experience with an assessment format might contribute to student success. Finally, it identified younger students’ perceptions of computer-based testing and focus on their test-taking behaviors. By conducting this study, literature in this field was be broadened, which is important because of the growing number of schools and students that are using computer-based testing.

CHAPTER III:

METHODS AND PROCEDURES

The study that was conducted was a quantitative study, which compared the growth scores of the students in their first year of taking the *i-Ready* Computer Adaptive Test with their growth scores during their second year of using the same testing program. In addition, the study compared the pilot group's scores the second year of testing with the group of students who is taking the test for the first year to indicate if experience with a test helps to improve student growth and achievement. A qualitative questionnaire was used for descriptive analysis in order to evaluate students after they have tried using the Sample Version of the New York State Test. This study investigated the following:

1. How does exposure and experience with a Computer Adaptive Test (the *i-Ready Diagnostic)* impact student growth when taken the first year compared to the student growth when taken the second year?
2. Is there a difference between the growth scores of the students who have had experience with the i-Ready Diagnostic and the student who have not had experience with the assessment program?
3. What effect does gender and instructional program have on student performance and student growth on a Computer Adaptive Test, such as the *i-Ready Diagnostic*?
4. What are students' perceptions of taking a computer-based state test at the elementary level in fourth and fifth grade after navigating through the sample exam provided by the New York State Education Department?

Hypotheses and Questions

Quantitative Hypothesis

Question 1

H_0: There is no difference in the growth scores from pre-assessment (January) to post-assessment (May) in English Language Arts between the 3rd and 4th grade students taking the *i-Ready* Computer Adaptive test in 2017-2018 and the same 4th and 5th grade students taking the test in 2018-2019.

H_1: There is a difference in the growth scores from pre-assessment (January) to post-assessment (May) in English Language Arts between the 3rd and 4th grade students taking the *i-Ready* Computer Adaptive test in 2017-2018 and the same 4th and 5th grade students taking the test in 2018-2019.

Question 2

H_0: There is no difference between the growth scores of the students who have had experience with the i-Ready Diagnostic and the students who have not had experience with the assessment program.

H_1: There is a difference between the growth scores of the students who have had experience with the i-Ready Diagnostic and the students who have not had experience with the assessment program.

Question 3

H_0: There is no difference in the growth scores on the *i-Ready* English Language Arts Diagnostic test between male and female students, nor students who are in different reading instructional programs between the years of 2017-2018 and 2018-2019.

H_1: There is a difference in the growth scores on the *i-Ready* English Language Arts Diagnostic test between male and female students, nor students who are in different reading instructional programs between the years of 2017-2018 and 2018-2019.

Descriptive Analysis

Question 4. What are students' perceptions of taking a computer-based state high-stakes test at the elementary level in fourth and fifth grade after navigating through the sample English Language Arts exam provided by the New York State Education Department?

Research Design and Data Analysis

The present study combined inferential and descriptive measures to provide a perspective on student use of computer-based assessments. The quantitative component of this study was an ex post facto design, as the data being collected does not impact or manipulate the participants and their participation in the diagnostic test taking. A multivariate analysis compared student growth scores in English Language Arts between two groups (experienced and not experienced with i-Ready assessments), across two grade levels (4th and 5th grade). Covariates of student performance included their reading scores, class grades, students that receive Academic Intervention Support and their experience with computers in the classroom, specifically experience with computer-based testing.

To examine if the assumptions of the design are met, a Levene's test was used to determine in the variances of the two populations are equal. To assess that the data set meet the parameters for multivariate analysis, skewness and kurtosis assessed the symmetry of the data plots. Skewness and kurtosis of the data set can be seen in Table 2. The distribution of student growth scores for year 1 can be found in Figure 2, and the distribution of growth scores for year 2 can be found in Figure 3. A power analysis determined the adequacy of the sample size, given the variables to be included.

Table 2

Descriptive Statistics: Skewness and Kurtosis

Statistic	Growth Scores Year 1	Growth Scores Year 2
n		
Valid	45	224
Missing	179	0
Skewness	.729	-.048
Std. Error of Skewness	.354	.163
Kurtosis	1.640	.511
Std. Error of Kurtosis	.695	.324

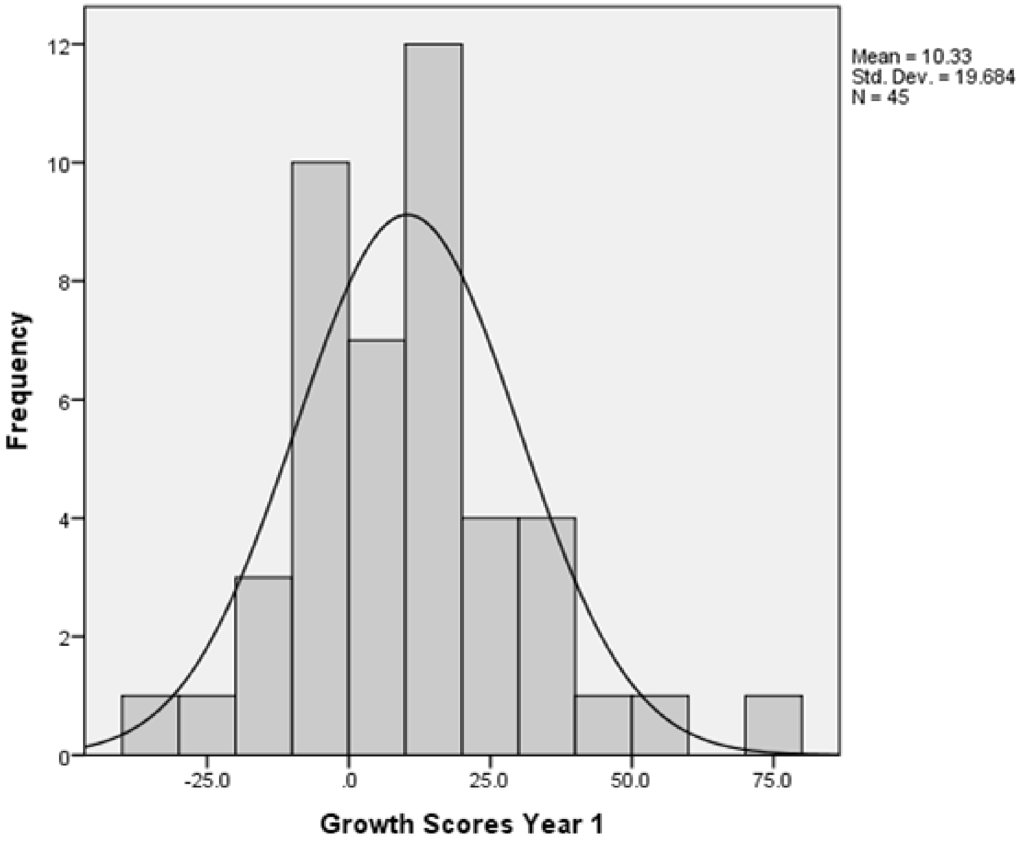


Figure 2. Distribution of scores in year 1.

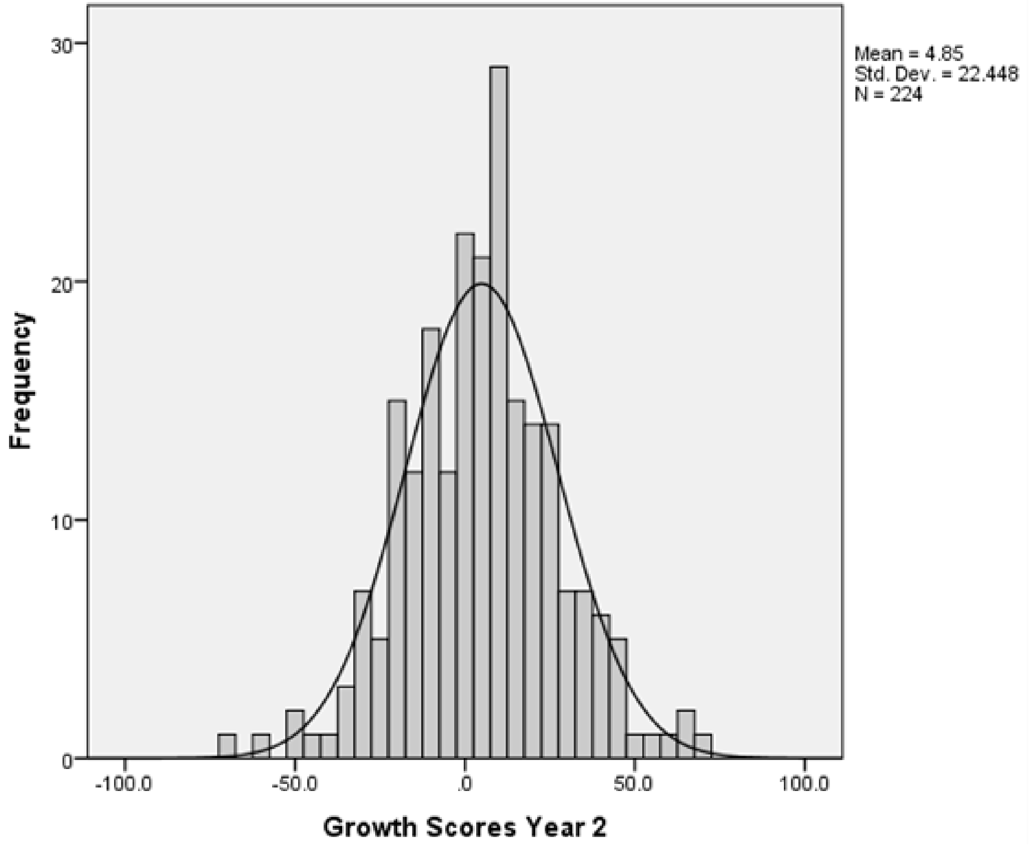


Figure 3. Distribution of scores in year 2.

The descriptive analysis portion of the study consisted of an open-ended questionnaire given to two classes of students (*n*=27) after they have completed an online assessment. The teacher-observers took notes on the following student behaviors while students are engaged in completing items on the practice New York State test (available from the NYSED website): time spent on reading the directions, interaction with the features on the online assessment, utilization of the features on the reading sample, and other behaviors, including looking for peer or teacher assistance, fidgeting, or rushing (clicking quickly) through the set of questions.

Population and Sample

Population

This study was conducted in a suburban school district of 51,881 in the northeastern United States. The school population is 86% white, 9% Hispanic, 4% Asian or Pacific Islander, and 1% other. All students in the school use the *Journeys* reading program, which was adopted in the district in 2012. In an effort to collect standardized data across the district, the district piloted the *i-Ready Computer-Based Reading Diagnostic* Assessment in 2017 and purchased the program for universal use in 2018. These Diagnostic tests are given three times throughout the school year.

Sample

A sample of 224 students from one school in this suburban district participated in the study. The data collected in this study was taken from the 3rd grade and the 4th grade who piloted the program in 2017-2018 (*n*=45). The study looked at their growth scores over two years. The growth scores from their first year of using the I-Ready ELA computer-adaptive assessment were compared to their growth scores from their second year using the same assessment program (2018-2019). In the second year of testing administration, the program was rolled out to the student body. The researcher collected the scores of students in 4th grade and 5th grade who were taking the test for the first time (*n*=179). The student scores of the pilot group were compared with the scores of the group of students taking the test for the first time. Student data gathered for this portion were anonymous. Parental permission was required for any student who participates in the questionnaire portion of the study.

Instruments

The *i-Ready Diagnostic* test for English Language Arts were used as the Computer Adaptive Test for the quantitative component of this study. According to Curriculum Associates (2018), the i-Ready test is reliable and valid. It was developed by "well-known experts in Educational Measurement, Computer Adaptive Testing, Mathematics, English Language Arts and the Common Core, adheres to the Standards of Psychological and Educational Testing and was independently audited for adherence to the Standards by researchers from the University of Massachusetts at Amherst…[and has] strong test metrics: Low SEMs; good item discrimination among students of different abilities. [Lastly, the test is] strongly correlated to Common Core assessments based on third-party research from the Educational Research Institute of America (ERIA)" (Curriculum Associates, 2018, p. 10).

The i-Ready is based on a raw score out of 800, which is based on the number of questions answered correctly versus the number of questions answered incorrectly. There is not a set amount of questions given to each student because the students' test items vary with each response. However, the test time usually last between 35 and 60 minutes. The i-Ready English Language Arts test is made up of six domains: Phonological Awareness, Phonics, High Frequency Words, Vocabulary, Comprehension of Literature, and Comprehension of Informational Text. Students in grades four and five often test out of the Phonological Awareness, Phonics, and High-Frequency Words domains, so their scores generally consist of the other three domains (Curriculum Associates, 2018).

Teacher-Observer Protocol

During the work time for the survey (one 40-minute period), two teachers observed student behavior, including how the students are interacting with the test, how well they are using navigation tools, and to what extent the students are using options such as highlighter and changing the color of the page. The two teachers were asked for feedback after their students take the practice test online. The researcher conducted a 40-minute training during the teachers' preparation periods, for delivering the questionnaires to the students and observing the students.

During the training session, the researcher provided the four teachers (two teachers for each survey session) with instructions for how to observe the students. The teachers were provided with an overview of the survey and online sampler. They were also shown how to navigate the New York State Education Department (NYSED) website, if a student were to have difficulty or click out of the website.

The teachers were taught how to use interval observations and were provided with stop watches, if requested. The observers were asked to stand behind each student and observe their behavior for two minutes. They were encouraged to use the checklist provided and take low-inference notes. After two minutes, the teachers were asked to move on to the next student for observations. By starting at opposite ends of the room and using the students' numbered computers, the observers were able to observe all students, meaning each student was observed two times.

A checklist of observable behaviors was used by the observers. The checklist is shown in Table 3. Before the observations took place, the teachers were trained by the researcher.

Table 3

Teacher-Observer Checklist

Student # (Based on computer station at which the student is working)	Behaviors used while being observed – check all that apply (during 2-minute interval)	Additional Comments or Anecdotal
	o Uses mouse to point to directions o Acknowledges none/some/all of the accommodations from the menu by clicking each icon. o Clicks "Continue" button without reading all directions. o Looks at other students' computers and/or moves eyes away from the computer screen regularly. o Follows the prompts carefully, as indicated by eyes focusing on the computer. o Student whispers what s/he is reading. o Student asks many questions or appears worried or overwhelmed.	

This information gathered by the observer was also be used in the descriptive analysis.

Questionnaire Protocol

The researcher modified the survey and interview questions used the study by Özden et al. (2004) to create questions better suited for the student participants at the elementary level. The ten questions were field tested by the researcher by giving the questions to ten students of the same age and asking two teachers to see if the questions were age appropriate. A blueprint of the student questions can be found in Table 4.

The questionnaires were given to the students by the teachers of each class. The students from grade 4 and grade 5 were asked the same set of questions (listed in the procedure section below). The questions given have been field tested by the researcher in order to ensure validity and reliability.

Procedures

In 2017-2018, grades 2-5 in this suburban school piloted the i-Ready Computer Adaptive Test in Reading in January and June, using one class from each grade. In 2018-2019, the school adopted the assessment tool for all classes in all grades and classes K-5 for September, January, and June. The researcher collected the *i-Ready* baseline data from the English Language Arts Diagnostic Tests from the students in grade 3 (*n*=45) from 2017-2018. The same data was collected for the same students in the baseline group, one year later, during the same time interval (January through June) in grades 4 and 5 to see if their experience after a year of using the program contributes to their growth over the course of the year (Table 5).

Table 4

Student Survey Questions and Connection to Theory and Literature

Student Questions	Connection to Theory and Literature
1. Have you used computer-based tests in school before now?	Student experience and readiness (Bruner, 1964, as cited by Schunk, 2016)
2. Was the computer screen easy to use when you took the sample test? Do you think that a tablet might be better?	Computer-Based Testing Format (Jarodzka et al., 2015)
3. Which of the tools did you use? Is the toolbox of this online assessment system easy to use?	Computer-Based Testing Format (Jarodzka et al., 2015)
4. Do you think that you using the computer for tests is more motivating than tests on paper? Explain.	Self-Regulated Learning, as it pertains to Self-Reflection (Zimmerman & Moylan, 2009)
5. What are the difficulties you faced while using the online assessment system?	Self-Regulated Learning, as it pertains to Self-Observation and Metacognitive Monitoring (Zimmerman & Moylan, 2009)
6. What did you like most while using the online assessment system?	Self-Regulated Learning, as it pertains to Self-Observation and Metacognitive Monitoring (Zimmerman & Moylan, 2009)
7. How would you make this computer test better or easier to use?	Computer-Based Testing Format (Jarodzka et al., 2015)
8. Was it helpful to practice using this sample test? Why or why not?	Student experience and readiness (Bruner, 1964, as cited by Schunk, 2016)
9. Do you think that the i-Ready test helped you to work this computer test sample?	Student experience and readiness (Bruner, 1964, as cited by Schunk, 2016)
10. When you finished, did you go back and check your work?	Self-Regulated Learning, as it pertains to Self-Reflection (Zimmerman & Moylan, 2009) -Testing Format Differences (Jeong, 2014)

The results of subgroups were analyzed to determine if there are significant differences based on gender or academic intervention services in reading.

Table 5

Quantitative Procedures for Data Collection

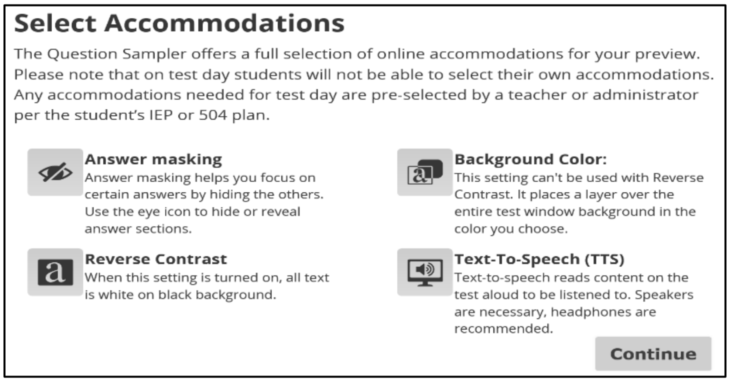
School Year	Grades	*n*	Procedure
2017-2018	3 and 4	45	Analyze growth score for the students between January, and June
2018-2019	4 and 5	224	Analyze growth scores for the students between January, and June. Use this data to compare: • the growth of the baseline group in their first year to their growth in the second year • the growth of the male students compared to the growth of the female students • the growth of the students receiving Reading Academic Intervention using the *i-Ready* Instructional Component five times per week for 42 minutes (*n*=10) compared to their peers who only use *i-Ready* for the Diagnostic Tests

Descriptive Analysis Procedure

The qualitative portion of the study included two classes of students: one in 4th grade and one in 5th grade. The students were asked to try and navigate through the practice, computer-based version of the New York State Test, available at nysed.gov (http://www.nysed.gov/edtech/question-sampler). The researcher provided the students with a class period during their school day, which is 40 minutes. This is the average time the students spend taking a test in the school, and it is the amount of time they are given in the computer lab on a weekly basis. They were asked to read through the directions and complete as much as they can.

Figures 4-7 are examples of pages that are shown on the test sampler. The observers made notes if students are exploring these options or if they are simply clicking “Continue” to begin the test. Figure 4 displays test accommodations that students may choose for their test, such as changing the contrast of the test or the background colors.

Figure 5 displays information about the test sampler's screen splitting capability. Figure 6 displays the tool options for the test sampler, and Figure 7 displays information about the navigation of the test sampler.



Select Accommodations

The Question Sampler offers a full selection of online accommodations for your preview. Please note that on test day students will not be able to select their own accommodations. Any accommodations needed for test day are pre-selected by a teacher or administrator per the student's IEP or 504 plan.

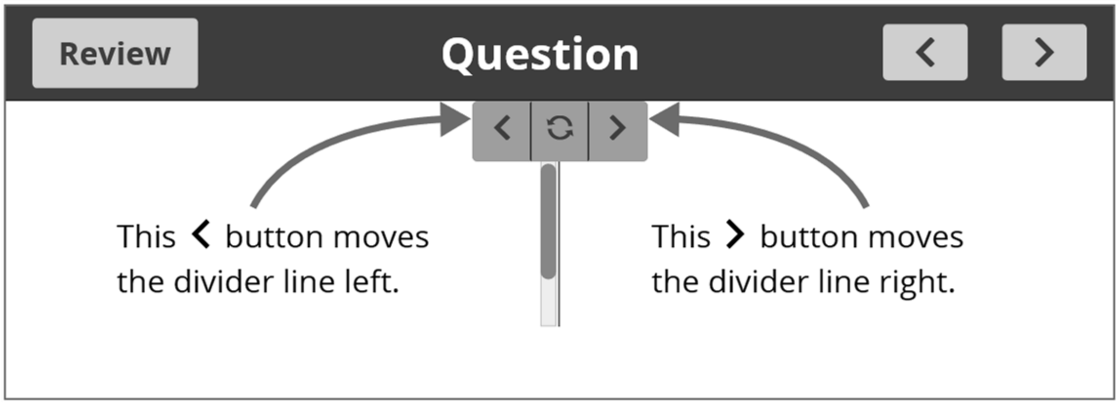
Answer masking
Answer masking helps you focus on certain answers by hiding the others. Use the eye icon to hide or reveal answer sections.

Background Color:
This setting can't be used with Reverse Contrast. It places a layer over the entire test window background in the color you choose.

Reverse Contrast
When this setting is turned on, all text is white on black background.

Text-To-Speech (TTS)
Text-to-speech reads content on the test aloud to be listened to. Speakers are necessary, headphones are recommended.

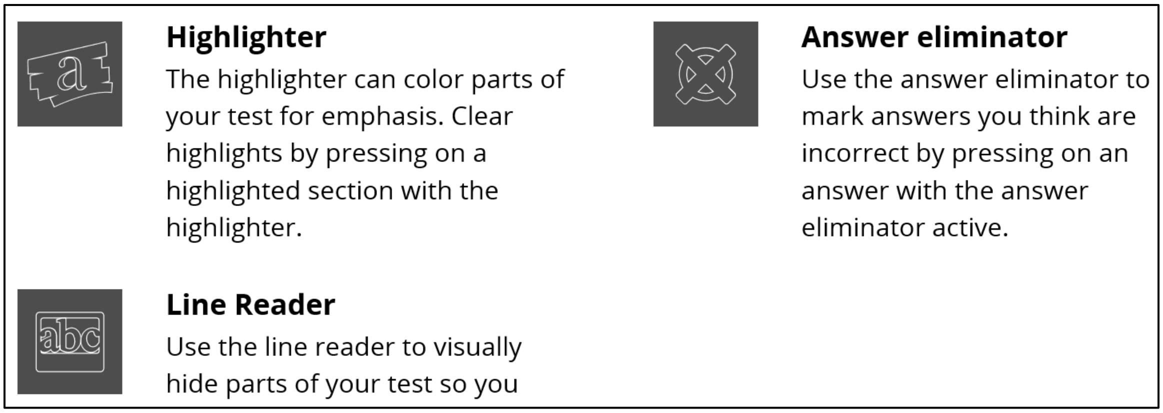
Figure 4. New York State ELA test sample accommodations, 2019. Retrieved from https://ny.nextera.questarai.com/tds/#practice.



This ❮ button moves the divider line left.

This ❯ button moves the divider line right.

Figure 5. New York State ELA test sample screen splitting tool, 2019. Retrieved from https://ny.nextera.questarai.com/tds/#practice.

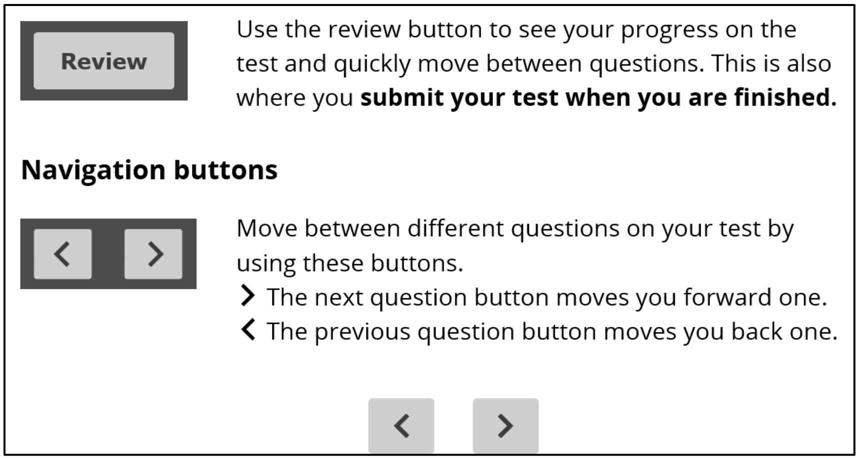


Highlighter
The highlighter can color parts of your test for emphasis. Clear highlights by pressing on a highlighted section with the highlighter.

Answer eliminator
Use the answer eliminator to mark answers you think are incorrect by pressing on an answer with the answer eliminator active.

Line Reader
Use the line reader to visually hide parts of your test so you

Figure 6. New York State ELA Test Sample Tool Options, 2019. Retrieved from https://ny.nextera.questarai.com/tds/#practice.



Use the review button to see your progress on the test and quickly move between questions. This is also where you **submit your test when you are finished.**

Navigation buttons

Move between different questions on your test by using these buttons.
> The next question button moves you forward one.
< The previous question button moves you back one.

Figure 7. New York State ELA test sample navigation tools, 2019. Retrieved from https://ny.nextera.questarai.com/tds/#practice.

Immediately following the sample test, an open-ended questionnaire was given, which asks the students about their computer testing experience. Responses were coded based on students’ responses and organized into categories including: Self-Regulated Behavior, Motivation, and Challenges.

CHAPTER IV:

RESULTS

Research was conducted to study the effect of experience with computer-based testing on student growth scores on the i-Ready diagnostic English Language Arts test. A group of students took the diagnostic test in January 2018 and May 2018 (*n*=45). The following school year, the i-Ready diagnostic was rolled out in September 2018, January 2019, and May 2019 (*n*=224). The researcher compared the diagnostic scores of the two groups from January to May 2019 to answer the following questions:

1. Were there differences from pre-assessment (January) to post-assessment (May) in English Language Arts between the 3rd and 4th grade students taking the *i-Ready* Computer Adaptive test in 2017-2018 and the same 4th and 5th grade students taking the test in 2018-2019?
2. Were there differences between the growth scores of the students who have had experience with the *i-Ready Diagnostic* and the students who have not have experience with the program?
3. What effect did gender and instructional program have on student performance and student growth on a Computer Adaptive Test, such as the *i-Ready Diagnostic*?

Question 1

Table 6

Paired Samples Statistics of Growth Scores for Pilot Group (Year 1 and Year 2)

	M	*n*	*SD*	Error
Growth Scores Year 1	10.333	45	19.6839	2.9343
Growth Scores Year 2	6.867	45	24.3054	3.6232

Table 7

Paired Sample Correlations of Growth Scores for Pilot Group (Year 1 and Year 2)

	n	Correlation	*p*
Growth Scores Year 1 Growth Scores Year 2	45	.274	.069

Table 8

Paired Sample t-Test of Growth Scores of Pilot Group (Year 1 and Year 2)

	Paired Differences					*t*	*df*	Sig. (2-tailed)
	M	*SD*	Error	95% Confidence Interval of the Difference				
				Lower	Upper			
Year 1 Year 2	3.47	26.77	3.99	-4.57	11.51	.869	44	.390

A paired sample t-Test was conducted to determine the effect of experience has on students' i-Ready ELA scores of the Pilot group. The test indicates that the difference in the mean of Growth Scores for year one (*n*=45, *M*=10.33, *SD*=19.68) and the mean Growth Scores for year two (*n*=45, *M*=6.87, *SD*=24.31) were not statistically significant, *t*(44)=3.47, *p*=.390.

Question 2

Table 9

Growth Scores of Students for 2018-2019 School Year

Group	*n*	*M*	*SD*	Error
Pilot Group	45	6.867	24.3054	3.6232
Full Roll Out	179	4.341	22.0000	1.6444

Table 10

Independent Sample t-Test Comparing Pilot Scores and Full Roll Out

	Levene's Test for Equality of Variances		t-test for Equality of Means						
	F	Sig.	*t*	*df*	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
								Lower	Upper
Growth Scores Jan-May									
Equal variances assumed	.795	.373	.674	222	.501	2.53	3.75	-4.86	9.91
Equal variances not assumed			.635	63.33	.528	2.53	3.98	-5.42	10.48

An independent sample t-test was conducted to determine the effect experience has on students' i-Ready growth scores. The test indicates that the difference in the mean of i-Ready growth scores for the students in the Pilot group (*n*=45, *M*=6.87, *SD*=24.31) and students in the full Roll Out (*n*=179, *M*=4.34, *SD*=22.00) were not statistically significant, *t*(222)=2.53, *p*=0.37.

Table 11

Analysis of Variance Between Pilot Group and Full Roll Out

Tests of Between-Subjects Effects						
Source	Type III Sum of Squares	*df*	Mean Square	*F*	*p*	Partial Eta Squared
Corrected Model	229.427[a]	1	229.427	.454	.501	.002
Intercept	4516.802	1	4516.802	8.941	.003	.039
Group	229.427	1	229.427	.454	.501	.002
Error	112145.412	222	505.160			
Total	117640.000	224				
Corrected Total	112374.839	223				

Note. [a] R Squared = .002 (Adjusted R Squared = -.002).

An analysis of variance showed that the effect of experience with the test was not significant for the growth scores on the i-Ready diagnostic, *F*(1,222) = .45, *p* = .501

Question 3

Table 12

Independent Sample t-Test to Compare Male and Female Growth Scores

	Levene's Test for Equality of Variances		t-test for Equality of Means						
	F	*p*	*t*	*df*	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
								Lower	Upper
Equal variances assumed	.105	.746	.227	222	.821	.6813	3.01	-5.24	6.61
Equal variances not assumed			.227	221.699	.821	.6813	3.00	-5.24	6.60

An independent sample t-Test was conducted to determine the effect gender has on students' i-Ready growth scores. The test indicates that the difference in the mean of i-Ready growth scores for the female students (*n*=111, *M*=4.51, *SD*=21.87) and male students (*n*=113, *M*=5.19, *SD*=23.10) were not statistically significant, *t*(222)=3.01, *p*=0.746.

Table 13

Analysis of Variance Between Instructional Groups and Gender (Year 1 and 2)

Source	Dependent Variable	Type III Sum of Squares	*df*	Mean Square	*F*	Sig.	Partial Eta Squared
Corrected Model	Year 1	264.38[a]	2	132.19	.331	.720	.016
	Year 2	1828.34[b]	2	914.17	1.589	.216	.070
Intercept	Year 1	2548.69	1	2548.69	6.378	.015	.132
	Year 2	3320.76	1	3320.76	5.772	.021	.121
Gender	Year 1	91.52	1	91.52	.229	.635	.005
	Year 2	205.93	1	205.93	.358	.553	.008
Instructional Group	Year 1	106.88	1	106.88	.267	.608	.006
	Year 2	1254.05	1	1254.05	2.180	.147	.049
Gender * Instructional Group	Year 1	.000	0	.	.	.	.000
	Year 2	.000	0	.	.	.	.000
Error	Year 1	16783.62	42	399.61			
	Year 2	24164.86	42	575.35			
Total	Year 1	21853.00	45				
	Year 2	28115.00	45				
Corrected Total	Year 1	17048.00	44				
	Year 2	25993.200	44				

Note. [a] R Squared = .016 (Adjusted R Squared = -.031); [b] R Squared = .070 (Adjusted R Squared = .026).

An analysis of variance was also conducted to investigate the effect of the instructional group and gender on student performance in ELA based on the i-Ready Diagnostic Growth scores. The results of the MANOVA are not significant when measuring student growth based on instructional group, *F*(1,42)=0.267, *p*=.147 and based on gender, *F*(1,42)=0.358, *p*=.553.

Table 14

Multivariate Tests

Effect		Value	*F*	Hypothesis *df*	Error *df*	*p*	Partial Eta Squared
Intercept	Wilks' Lambda	.812	4.735[b]	2.000	41.000	.014	.188
Gender	Wilks' Lambda	.989	.231[b]	2.000	41.000	.795	.011
Instructional Group	Wilks' Lambda	.950	1.075[b]	2.000	41.000	.351	.050
Gender * Instructional Group	Wilks' Lambda	1.000	.[b]	.000	41.500	.	.

Note. [a] Design: Intercept + Gender + InstructionalGroup + Gender * InstructionalGroup.

Table 15

Growth Scores Based on Gender and Instructional Group, Year 1 and 2

	Gender	Instructional Group	*M*	*SD*	*N*
Growth Scores Year 1	Female	Tier 1 or 2	8.33	22.45	21
		Total	8.33	22.45	21
	Male	Tier 1 or 2	11.29	18.26	21
		i-Ready	17.67	4.04	3
		Total	12.08	17.21	24
	Total	Tier 1 or 2	9.81	20.27	42
		i-Ready	17.67	4.04	3
		Total	10.33	19.68	45
Growth Scores Year 2	Female	Tier 1 or 2	3.05	21.21	21
		Total	3.05	21.21	21
	Male	Tier 1 or 2	7.48	24.16	21
		i-Ready	29.33	41.79	3
		Total	10.21	26.72	24
	Total	Tier 1 or 2	5.26	22.57	42
		i-Ready	29.33	41.79	3
		Total	6.87	24.31	45

While there were no significant differences in growth between the instructional groups, it is valuable to note that the scores of the students in the i-Ready Instructional Group were higher than those in the Tier 1 and Tier 2 groups. It should be noted that they i-Ready Instructional Group was used in the sample, but has a very low sample size. Accordingly, there is no significant difference between the mean scores of the males and females, it is noteworthy that for both genders the growth scores decreased from year one to year two. However, the males' scores were higher than females' scores both years.

Descriptive Statistics

A survey was conducted by the researcher after the students took an English Language Arts Test Sampler that used a different format from i-Ready diagnostic. The survey examines students' motivation and perceptions and answers the question:

> *What are students' perceptions of taking a computer-based state high-stakes test at the elementary level in fourth and fifth grade after navigating through the sample English Language Arts exam provided by the New York State Education Department?*

Table 16

Student Survey Participants

Survey Participants	Girls (*n*)	Boys (*n*)	Total (*n*)
Grade 4	5	8	13
Grade 5	9	5	14

A survey was conducted after 4th grade students (*n*=13) and 5th grade students (*n*=14) took the New York State ELA Sample Test. All students were given 20 minutes to complete the sample test provided by *Questar*, which included a reading passage accompanied by five comprehension questions (four multiple choice, one written response). After 20 minutes, each student took a 10-question survey.

Table 17 includes the responses from the survey.

Table 17

Student Survey Responses

Student Questions	Student Responses
Have you used computer-based tests in school before now?	26 students responded yes that they have taken the i-ready test. No other tests were mentioned.
Was the computer screen easy to use when you took the sample test? Do you think that a tablet might be better?	26 students said that it was easy to use 14 students said that they would prefer a tablet because it would be easier to use (easier to scroll, highlight, click)
Which of the tools did you use? Is the toolbox of this online assessment system easy to use?	High lighter -9 Line-reader- 6 Answer eliminator- 4 Note taker=3 Zoom- 2 None -12
Do you think that you using the computer for tests is more motivating than tests on paper? Explain.	Prefer Computer- 14 Prefer paper- 6 Unsure- 2
What are the difficulties you faced while using the online assessment system?	How to use the tools- 7 No difficulties- 7 Moving to the next page/ Navigation/scrolling/mouse- 13
What did you like most while using the online assessment system?	Using the tools- 10 Screen Splitting- 3 Using the computer (typing answers, clicking on questions, no paper)- 12 Didn't like anything - 1
How would you make this computer test better or easier to use?	Navigation (scrolling and going to "next" page)- 7 Screen size - 2 Make highlighter easier to use – 3 Make the directions easier to understand (ie, tools tutorial, instead of labeled directions)- 5 Change nothing- 5 Miscellaneous- 7
Was it helpful to practice using this sample test? Why or why not?	Yes- 21 No- 3 Unsure- 3
Do you think that the i-Ready test helped you to work this computer test sample?	Yes- 12 No- 13 Unsure- 2
When you finished, did you go back and check your work?	No - 8 Yes- 17 I didn't know how to- 2

These results indicate that students did have experience, and most felt that the test was easy to use. However, more than half of the students surveyed did suggest that a tablet would be better to use than a computer. Accordingly, many of the students indicated that they faced difficulties with navigating through the test. Some students cited that scrolling through the screen was difficult, while others stated that they did not know how to move on to the “next page”.

According to Question 4, most students felt motivated by using the computer and indicated that they prefer using the computer over paper. In fact, according to Question 6, when asked what they liked most about the computer-based test, 12 said that they enjoyed using the computer, typing, and being able to click their answers; 10 students liked the tools; 3 students indicated that they liked the screen splitting; and only one student indicated that s/he did not enjoy anything about the computer-based test.

According to Question 9, 21 students felt that using this test was valuable and helpful. However, only 12 students indicated that the i-Ready helped to prepare them for the test that they took; 13 students said that it the i-Ready did not help them, many of them indicating that the two formats were different. This indicates that testing format and format consistency may be useful when computer-based tests are developed.

Observers’ Notes

During the qualitative portion of the study, two teacher observers were trained in order to take low-inference notes on student behaviors during the New York State Test Sampler. Their observations are shown in Table 18.

Table 18

Observer Questionnaire Results

Student Test-Taking Behavior	Number of Students
Uses mouse to point to directions	11
Acknowledge some/all of the accommodations by clicking the icons	5
Clicks "Continue" button without reading all directions	16
Looks at other students' computers and/or moves eyes away from computer screen regularly	6
Follows the prompts carefully, as indicated by eyes focusing on the computer	5
Student whispers what s/he is reading	4
Student asks many questions and/or appears worried or overwhelmed	7

Other student behaviors noted by the anecdotal comments made by the observers included:

- Twelve students did not use any tools
- The most used tool by the students was the highlighter
- More than half of the students had trouble navigating the screen, specifically how to move to the next question (because the screen splitting tool and the "next page" command were two arrows that looked similar)
- Two students who used the note-taker did not know how to minimize their notes and retrieve them when needed
- When the students were told that five minutes were left, the observers were asked to note if students went back to check their work. The observers noticed that only five students went back to check their work.

The data collected by the teacher observers allows educators to examine areas in which they might need to more explicitly and carefully present test-taking strategies to students in their classes. In this case, it was clear that many students did not acknowledge all of the information presented in the directions, that navigation was a concern for both teachers and students, and that the main “tool” used by students was simply the mouse in order to track the words on the screen.

Conclusion

The results of the present study provide some promising support for the use of CBT with elementary age students in terms of student ability to complete the tasks and absence of significant gender differences. Issues of self-regulation of young students must be considered, however, based on overall student performance. Further, the descriptive analyses raise concerns about use of the tools provided to students, as well as student understanding and motivation when tests are presented online. The implications of the data analyses are discussed in the next chapter.

CHAPTER V:

DISCUSSION

In this study, a pilot group of students took the *i-Ready Diagnostic* exam in 2017-2018. Then, the *i-Ready* assessment was rolled out to the rest of the student body. The results of the two years were recorded in order to analyze the student growth in English Language Arts to see if:

1. There was a difference in student scores based on experience with the *i-Ready* test
2. There was a difference in student scores based on gender or reading instructional program

After the quantitative portion of the study, a qualitative analysis was conducted using a survey to help identify student perceptions of computer-based tests.

This chapter reviews the data presented in Chapter IV and connect it to the literature and theoretical framework. The findings and data helped to make some recommendations to administrators and professionals in the field of education and assessment, to help them make decisions about types of assessments that they choose to use in the future.

Implications of Findings

The findings of this study indicate that there were no significant differences of student growth between students with experience and students without experience on the *i-Ready* assessments. This may indicate that the students in grades 3 through 5 are not equipped with the self-regulatory behaviors that are needed in order to be successful on adaptive tests. According to Zimmerman and Moylan (2009), self-regulation includes not only strategy and time management, but also self-consequences and metacognitive monitoring. After the tests, students should exhibit self-judgement, which can include goal setting and planning for future assessments (Panadero, 2017). If there were no significant differences between students who had experience and those that did not, it might be assumed that self-regulatory behaviors might need to be taught more explicitly, so that they can better apply the skills to the new testing formats, such as computer-adaptive tests.

This lack of self-regulation was also seen when the students were taking their surveys. On student surveys, 17 students (out of 27) indicated that they checked their work. However, the adult observers only recorded that 5 students went back to check their work. This reveals that students may misunderstand what it means to check work, or they might need to be explicitly taught how to review their work before handing it in. From this portion of the test, it was clear that student participants needed to better develop their self-regulatory behavior with help from their teachers because in order to find success and growth on the adaptive tests, reflection and goal setting is important to future successes.

Relationship to Prior Research

The findings of this study led to the acceptance of a null hypothesis, in which there were no significant differences between test scores over time nor between groups of students based on instructional group or gender. Results of this study indicate that student achievement on the i-Ready diagnostic test did not vary significantly for students that had experience with the computer-based test after two years. The scores were also not significantly different between groups of students, based on gender or instructional groups.

The Effect of Experience

The results of the study conducted by Backes and Cowan (2019) indicate that in English Language Arts, testing administration did account for a portion of the mode effect, as did student experience. However, this study found that there was no significant difference between the scores of the students with experience and without experience with the *i-Ready Diagnostic* test.

Because there were no significant differences in student achievement during this study, teachers may want to consider more explicit instruction. For example, in a study conducted by Worrell et al. (2016), four students were systematically taught the "NRUN" mnemonic strategy to help them better perform on computer-based tests. The purpose of the study was to see if the reading strategy NRUN would be used by students when interacting with the text on a computer. With explicit instruction of the reading strategy, the students' computer-based test scores increased. While the sample size of Worrell et al. (2016) is small it may encourage teachers to attempt teaching test-taking strategies in the future in order to increase *i-Ready* assessment scores.

More research needs to be conducted in the area of experience with computer-based testing, as Backes and Cowan (2019) also indicated that even if experience correlated with student improvement, students who took paper-based versions of the PARCC exam still performed better than those students who took the computer-based test.

The Effect of Gender and Instructional Group

This study reported that the effect of gender and instructional group was not significant on the *i-Ready Diagnostic* test. Boys and girls in grades 3 through 5 did not differ significantly in their academic performances based on their gender nor their instructional grouping. This is different from the results of the study conducted by Jeong (2014), who found that the female participants' performance on the computer-based tests yielded significantly different scores in math, science, and Korean compared to the paper-based versions, whereas the difference in male scores were not significant. These contradictory results indicate that more research on the effect of gender is needed when implementing computer-based tests.

Student Perceptions

Özden et al. (2004) concluded that the key to student perception of online assessments is experience. However, experience did not affect student growth scores in the quantitative portion of the study. In order to take a closer look at the quantitative results, a descriptive analysis was conducted to see what students were thinking about as they took computer-based tests.

Similar to the results of Richardson et al. (2002), who reported that of the 24 student participants, 21 of them indicated that they preferred using the computer-based portions of the test, the survey used in the descriptive analysis portion of this study indicate that 14 students out of 27 found the Computer Based test to be motivating. Additionally, according to Richardson et al. (2002), students preferred the colors and images on the computer, the ease of typing (as opposed to an aching hand after writing on a paper-based test), and the tasks on the computer, which students said were more interesting than the paper-based tasks (p. 642). In the survey conducted for this study, students answered that they enjoyed using the tools and the ease of typing and clicking answers, rather than using a traditional paper-based test.

From these results, it is clear that most students enjoy the computer-based format, even if they have not performed well using such assessments.

Limitations of the Study

With the ex post facto design, there are many possible limitations, including threats to internal and external validity.

Threats to External Validity

There might have been interaction of testing and treatment, due to the repeated nature of the *i-Ready* Diagnostic. A limitation also includes interaction of setting and treatment because test-delivery may impact results. While the test is done on the computer, it is important to have an active proctor to help keep students on task. Additionally, the test time given for the students was 90 minutes. This was enough time for most students, however, some student did not finish and worked through the test in days that followed. Time also play a factor because having a 3rd or 4th grader take a test for more than an hour can cause testing fatigue and limit the self-monitoring skills after a certain amount of time.

Threats to Internal Validity

Two threat to internal validity include maturation, or the effect that passing time, resulting in growing older or more experienced, and testing, which may lead to students becoming familiar with the test. As with all school settings, there are many factors that affect academic achievement. For example, teachers have a great impact on students and their academic improvement. Therefore, growth in one year can differ from the following year with a different teacher. In this study, the teachers changed from year one to year two, so student growth on the test may have to do with comfort with the test format, as well as another outside factor.

Instrumentation is another limitation. The researcher has no control of the *i-Ready* adaptive test, and all tests are different. Therefore, changes in calibration or in the program over a two-year period were not accounted for in the research. Additionally, the scores that were calculated were very inconsistent. Over the five score intervals, the observer noticed that the scores were changing dramatically. Despite the *i-Ready's* claim that the assessment has been tested for "strong test metrics: Low SEMs; good item discrimination among students of different abilities. [Lastly, the test is] strongly correlated to Common Core assessments based on third-party research from the Educational Research Institute of America (ERIA)" (Curriculum Associates, 2018, p. 10), these large positive and negative swings in scores call into question the testing reliability. The reliability is not supported by the inconsistent student performance.

Lastly, mortality may play a role and impact statistical power if students leave the school or are absent during the week of testing.

Descriptive Analysis Limitations

As with the qualitative portion of the study, there are limitations regarding bias and interpretation. There may be a threat to descriptive validity during the observation portion, as the teachers may be unable to record all student behaviors. This may be coupled with interpretation validity and researcher bias, as the researcher may misinterpret or misconstrue the gathered data. Finally, the participants' reactivity may be a threat, as the students may change their behavior because they are being observed.

Students may also answer the questionnaire differently than normal due to the change in their computer class routine and their environment.

The information gathered by the student survey and the information gathered by the teacher observers did not match in one specific area. In part of the survey about checking work, 17 students indicated that they checked their work, but the observers only indicated that 5 students checked their work. In order to make this more accurate, the researcher should more clearly define what it means to "check work" for the students taking the surveys.

Recommendations for Future Practice

Despite resulting in a null hypothesis, this study offers educators, administrators and test-developers valuable lessons and recommendations.

In the current school climate, assessment scores and test scores are used as important tools for both teachers and administrators. Teachers use the scores for grouping students and providing parents with information about student progress. Administrators use test scores for rating schools within a district, as well as rating teachers. The *i-Ready* Diagnostic scores were very inconsistent throughout the entire sample. Student scores were often highest during their first tests, and then went up and down drastically as the year progressed. With the inconsistent score pattern educators should be cautious when using scores to determine student growth and teacher effectiveness.

To continue, more explicit instruction is needed for students to feel more confident when using a computer-based test. For example, students may need to use various assessments, in order to see tools that are consistently offered, such as the highlighting tools, screen splitting, and commonly used navigation symbols. Some students in the study used the note-taker, which is helpful, but only if students are proficient at typing. More time in computer labs or more time with 1:1 devices would help students to use computer-based assessments with ease.

Not only would it help for students to spend time with the computer-based tools, but it is also important for teachers to provide self-regulatory skills for students that would help students grow academically on and off the computer. For example, using mnemonic devices to help with reading comprehension, such as the NRUN device, can help to improve reading comprehension scores on and off the computer. Encouraging goal setting, self-evaluation, and checking over student work might also contribute to student success, especially when working on adaptive testing.

Based on the data, experience with the *i-Ready Diagnostic* did not help to improve test scores. One variable that was not examined was how many student evaluations do these students take each year. It may be helpful to limit the amount of testing used in a school. This was when students are taking an important assessment, they are putting their best effort into it. It is important for educators and administrators to be cautious of over-testing in schools at such young ages.

Lastly, test developers must consider engaging ways to deliver information about navigating the test and using the tools. As indicated by the surveys and observations, the students struggled with efficiently using the tools and moving from page to page. It was also noted that most students clicked quickly through the directions. Test developers need to consider ways encourage students to sit through the directions and tutorial. It might also be helpful to make the directions a guided audio and visual presentation, rather than having elementary aged students read and click through the directions on their own.

Recommendations for Future Research

The current study examined the student achievement of students using computer-adaptive computer-based testing. The study explored the differences between test scores based on experience, gender, and instructional reading programs. There were no significant differences found, but this study leads us to many more unexamined areas of assessment.

Based on the research provided in Chapter II, more research has been conducted at the high school, college, and post graduate than at the elementary level in the area of computer-based assessment. However, according to Backes and Cowan (2019), "Computer-based testing is rapidly spreading across the assessment landscape" (p. 89). More research is needed at the elementary level, in order for our schools and our students to be prepared for the inevitable changes in assessment.

Future research should explore the effect of computer-based testing in the various curriculum areas *and* the age at which they begin to test using computer-based assessments. Because self-regulation plays a role in students' success on assessment, it may be important to study different formats of computer-based tests that help to positively reinforce student progress with feedback or with an academic game. The devices used to assess students may be an area for potential research, in order to consider if the use of computers *or* tablets impact a students' performance.

Another area that requires more focus is classroom instruction. It may be helpful to explore how an increase in explicit classroom instruction on computer-based testing strategies helps to improve student performance. By having teachers spend time teaching self-regulatory strategies, as well as teaching basic computer skills, such as using appropriate tools, student scores may be impacted. While experience with the *i-Ready* diagnostic did not have significant effects on student scores, more regular use of computers and more explicit instruction on test-taking skills might make a difference.

To accompany the idea of studying more explicit instruction, it may be valuable to research how teachers are responding to computer-based programs and assessments. Using qualitative research, it would be helpful to investigate teacher perceptions, as well as best-practices for transitioning students from paper-based to computer-based testing. It is valuable to find out about the perceived obstacles that teachers are facing. By looking into teachers' perspectives, we may find other areas of professional development that need to be addressed in order for students to find success.

Another area of interest that one might explore is comparing student growth scores between schools with 1:1 devices and schools without. More regular and consistent use of instructional materials and assessments on the computer might help students who have 1:1 devices to perform better than peers in schools who only have access to computer labs on a weekly basis.

The final area of research that should be looked at it similar to the study conducted by Jeong (2014). Using standards-based testing (not adaptive testing), it would be helpful to see the achievement of elementary-aged students on computer-based and paper-based testing. This study focused on adaptive testing, so we were only comparing the growth, as all student received a different set of questions. Using a standard-based test, one can compare student achievement on the same computer-based and paper-based test. Backes and Cowan (2019) studied this using the PARCC exam roll out in Massachusetts. They found that there were many test mode effects that impacted student performance, that are not related to student ability (Backes and Cowan, 2019, p. 101). However, more research in this area need to be conducted. Student age, experience, socioeconomic status, and even testing administration by the school can impact how students are performing on computer-based tests. Comparing scores on the computer to the control of paper-based testing would help educators and policy-makers make more well-informed decisions about which assessments are reliable for students at the elementary level.

Computer-based testing is an important area of study for researchers in education because we are rapidly adopting more technology in schools each year. More research in this area will help educators and administrators adjust to the needs of the students who are taking the tests.

Conclusion

The use of computers in the classroom has increased over the past decade, and so too will the implementation of computer-based assessments. Educators must consider the programs that they are using and decide if they are valid and reliable for assessing their students. If a program is valid and reliable, then educators must better-prepare their students to take such assessments, by providing self-regulatory and test-taking strategies in order to help their students grow. As the transition from paper-based to computer-based assessment moves forward, administrators and policy makers need to allow schools time to adjust before using such tests as high stakes assessments and using them for teacher and school evaluations. Instead, schools, administrators, and policymakers need to work together to make the transition as smooth as possible.

This study found that one year of testing experience did not affect the student achievement compared to students with no experience. However, more experience and explicit instruction is needed for students, especially at the elementary level. Practice with computer-based assessments and instructional time that focuses on test-taking and self-regulation strategies is needed in the roll-out of such assessments.

More research needs to be conducted in this area. Administrators need to analyze the needs of their schools and better prepare their students for the computer-based assessments and the rise of 21st century learning skills.

REFERENCES

Backes, B., & Cowan, J. (2019). Is the pen mightier than the keyboard? The effect of online testing on measured student achievement. *Economics of Education Review, 68*, 89–103. doi:10.1016/j.econedurev.2018.12.007

Boevé, A. J., Meijer, R. R., Albers, C. J., Beetsma, Y., & Bosker, R. J. (2015). Introducing computer-based testing in high-stakes exams in higher education: Results of a field experiment. *Plos ONE*, *10*(12), 1-13. doi:10.1371/journal.pone.0143616

Brody, L. (2018, April 11). Computer glitches prevent some New York students from taking exams; This week marks the first widespread rollout of computer-based testing for grades three through eight. *Wall Street Journal.* Retrieved from https://jerome.stjohns.edu:81/login?url= ?url=https://search-proquest-com.jerome.stjohns.edu/docview/2023826333?accountid=14068

Burman, J., & Beattie, J. (2016). *New York State Education Department announces opportunities for computer-based testing for grades 3-8 ELA and math in 2017*. New York State Education Department. Retrieved from http://www.nysed.gov/news/2016/new-yorkstate-education-department-announces-opportunities-computer-based-testing-grades

Chang, C.-S. (2008). Development and validation of the computer technology literacy self assessment scale for Taiwanese elementary school students. *Adolescence*, *43*(171), 623–634. Retrieved from https://jerome.stjohns.edu/login?url=https://search.ebscohost.com/login.aspx?direct=tru&db=aph&AN=35390855&site=ehost-live

Chua, Y. P. (2012). Effects of computer-based testing on test performance and testing motivation. *Computers in Human Behavior, 28*, 5, 1580-1586.

Curriculum Associates. (2018). *The science behind the i-ready diagnostic*. Retrieved from https://www.curriculumassociates.com/Research-and-Efficacy.

Dixson, D. D., & Worrell, F. C. (2016). Formative and summative assessment in the classroom. *Theory Into Practice*, *55*(2), 153–159. doi:10.1080/00405841.2016.114898

Grant, M., Tamim, S., Brown, D., Sweeney, J., Ferguson, F., & Jones, L. (2015). Teaching and learning with mobile computing devices: Case study in K-12 classrooms. *Techtrends: Linking Research Practice to Improve Learning*, *59*(4), 32-45. doi:10.1007/s11528-015-0869-3

Greene, J. A., Moos, D. C., & Azevedo, R. (2011). Self-regulation of learning with computer based learning environments. *New Directions for Teaching & Learning*, *2011*(126), 107-115. doi:10.1002/tl.449

Herald, B. (2016). Seven studies comparing paper and computer test scores. *Education Week*, *35*(22), 8. Retrieved from https://jerome.stjohns.edu/login?url=https://search.ebscohost.com/login.aspx?direct=true&db=aph&AN=113413684&site=ehost-live

International Society of Technology Education [ISTE]. (2019). *Standards for students*. Retrieved from https://www.iste.org/standards/for-students.

Jarodzka, H., Janssen, N., Kirschner, P. A., & Erkens, G. (2015). Avoiding split attention in computer-based testing: Is neglecting additional information facilitative?

British Journal of Educational Technology, *46*(4), 803-817. doi:10.1111/bjet.12174

Jeong, H. (2014). A comparative study of scores on computer-based tests and paper-based tests. *Behaviour & Information Technology*, *33*(4), 410-422. doi:10.1080/0144929X.2012.710647

Merchant, G. J. (2004). What is at stake with high stakes testing? A discussion of issues and research. *Ohio Journal of Science, 104*(2), 2–7. Retrieved from https://jerome.stjohns.edu/login?url=https://search.ebscohost.com/login.aspx?direct=true&db=aph&AN=13590666&site=ehost-live

New York State Education Departments, 2019, www.nysed.gov.

Özden, Y. M., Ertürk, I., & Sanli, R. (2004). Students' perceptions of online assessment: A case study. *Journal of Distance Education, 19*(2), 77-94.

Panadero, E. (2017). A review of self-regulated learning: Six models and four directions for research. *Frontiers in Psychology*, *8*, 422. doi:10.3389/fpsyg.2017.00422

Pittman, T., & Gaines, T. (2015). Technology integration in third, fourth and fifth grade classrooms in a Florida school district. *Educational Technology Research & Development*, *63*(4), 539-554. doi:10.1007/s11423-015-9391-8

Richardson, M., Baird, J., Ridgway, J., Ripley, M., Shorrocks-Taylor, D., & Swan, M. (2002). Challenging minds? Students' perceptions of computer-based World Class Tests of problem solving. *Computers in Human Behavior*, *18*(6), 633.

Schunk, D. H. (2016). *Learning theories: An educational perspective* (7th ed.). Boston, MA: Pearson.

Shapiro, E. S., & Gebhardt, S. N. (2012). Comparing computer-adaptive and curriculum-based measurement methods of assessment. *School Psychology Review*, *41*(3), 295-305.

Sparks, S. (2015, November 11). Types of assessments: A head-to-head comparison. *Education Week*, *35*(12), s3. Retrieved from https://www.edweek.org/ew/section/multimedia/types-of-assessments-a-head-to-head-comparison.html

Worrell, J., Duffy, M. L., Brady, M. P., Dukes, C., & Gonzalez-DeHass, A. (2016). Training and generalization effects of a reading comprehension learning strategy on computer and paper-pencil assessments. *Preventing School Failure, 60*(4), 267–277. doi:10.1080/1045988X.2015.1116430

Zhang, M., Trussell, R., Gallegos, B., & Asam, R. (2015). Using math apps for improving student learning: An exploratory study in an inclusive fourth grade classroom. *Techtrends: Linking Research & Practice to Improve Learning*, *59*(2), 32-39. doi:10.1007/s11528-015-0837-y

Vita

Name	*Brittany A. Neligan*
Baccalaureate Degree	*Bachelor of Science*, *Bucknell University, Lewisburg, PA* *Major: Elementary Education*
Date Graduated	*May, 2008*
Other Degrees and Certificates	*Master of Science, City University of New York, Queens, NY* *Major: Special Education*
Date Graduated	*May, 2010*