2021

Influence of Istation Formative Assessment on Reading Student Achievement

Tiffini Ann Montelione
Walden University

Walden University

College of Education

This is to certify that the doctoral study by

Tiffini Ann Amato Montelione

has been found to be complete and satisfactory in all respects,
and that any and all revisions required by
the review committee have been made.

Review Committee
Dr. Michael Vinella, Committee Chairperson, Education Faculty
Dr. Anissa Harris, Committee Member, Education Faculty
Dr. Shereeza Mohammed, University Reviewer, Education Faculty

Chief Academic Officer and Provost
Sue Subocz, Ph.D.

Walden University
2021

Abstract

Influence of Istation Formative Assessment on Reading Student Achievement

by

Tiffini Ann Amato Montelione

MS, Dominican College of Blauvelt, 2004

BS, State University of New York at New Paltz, 2001

Dissertation Submitted in Partial Fulfillment

of the Requirements for the Degree of

Doctor of Education

Walden University

January 2021

Abstract

Many elementary teachers use formative assessment strategies as common practice, yet little information is available regarding specific programs and their influence on student achievement. Accordingly, the problem of this quantitative study was that the influence of the Istation computer-based reading program on student achievement as measured by the English Language Arts Partnership for Assessment of Readiness for College and Careers (ELA PARCC) summative assessment is unknown. The purpose was to investigate the influence of the Istation program on achievement and the use of specific digital learning tools for assessment. Sadler's framework outlined the formative assessment cycle, designed to improve achievement, and guided the research questions of how performance on Istation influenced summative assessment. The 6 research questions for the study investigated the differences in Istation's Indicators of Progress (ISIP) scores over time and the relationship between the Istation formative and ELA PARCC summative assessments. Ex-post facto analyses included a one-way repeated measures mixed analysis of variance and a linear regression analysis from a data sample from 175 Grade 3-5 students who each had a complete data set of 3 formative assessment ISIP scores from Istation and 1 summative PARCC score. Key results indicated ISIP scores improved over time and Istation can be used to predict student achievement on the ELA PARCC. Findings reinforced the potential of the program to offer student feedback and teacher data to guide instruction. Such a program can positively impact social change by improving essential student ELA skills and performance on the standardized state test that will translate to greater success in the future.

Influence of Istation Formative Assessment on Reading Student Achievement

by

Tiffini Ann Amato Montelione

MS, Dominican College of Blauvelt, 2004

BS, State University of New York at New Paltz, 2001

Dissertation Submitted in Partial Fulfillment

of the Requirements for the Degree of

Doctor of Education

Walden University

January 2021

Dedication

This study is wholeheartedly dedicated to my amazingly supportive and incredibly intelligent husband. Ryan: without you, I would surely have missed several meals and caffeinated drinks, and your belief in me kept me motivated beyond measure. You are the best partner in crime and life. I would also like to dedicate my dissertation in memoriam to those who did not make it to the end of this journey with me. To Nana: you were one of the strongest women I have ever known, and I would consider myself lucky if I were one day remembered for just half of your wisdom, humor, and kindness. My Della: I picture you laying here next to me every time I sit down to write, and you will be in my heart representing unconditional love forever. We did it!

Acknowledgments

I would like to express my deep and sincere gratitude to my committee members for their ongoing feedback and support. I could not have completed this dissertation without them. Dr. Vinella: your knowledge and demeanor helped me push through the most challenging times and, more often than I care to admit, sentences that did not even make sense to me. I am grateful that you were my chair. Dr. Harris: you will always hold a special place in my heart as the methodologist extraordinaire. Thank you for the early-morning and late-night text conversations, constant encouragement, and for helping me understand how to interpret ANOVA, regression, and beyond. Dr. Mohammed: thank you for your feedback throughout the entire University Research Review process.

A big piece of this work goes to my family and friends who gave me the confidence boosts, free editing services, hysterical laughing bouts, or shoulders to lean on when I needed these things the most. To my husband Ryan: thank you for holding down the fort, walking our girls, and for getting comfortable with statistical analysis conversations. To Mommy: thank you for always asking if I have eaten, for instilling the thought that I can do anything I put my mind to, and most importantly for being the epitome of perseverance and courage. To Jim: thank you for your encouragement and for handling the plight of teenage-me so I could make it to this point. To Daddy: thank you for telling me I was perfect from day one, for passing on your chill trait, and for your offerings of love and laughs along the way. Finally, to my girl crew—Victoria, Dianne, Corey, and Briele: I am forever grateful for our happy hours and your sustaining friendship with the hyper-focused Tiffini that made it through and achieved this goal.

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction to the Study

When executed correctly, formative assessments may inform teachers and their students about performance and skill progress. Sadler (1989) indicated that designing instruction in accordance with the formative assessment process increases student achievement on summative assessments. The Istation program is a formative assessment that teachers use to assess students at Hagan Elementary School (HES, pseudonym), the local district where this study took place. The research problem addressed in this dissertation is that the influence of the Istation reading formative assessment program on student achievement as measured by the English Language Arts (ELA) Partnership for Assessment of Readiness for College and Careers (PARCC) summative assessment is unknown. The study was conducted to fill the gap in practice regarding the link between specific formative assessments and summative assessment. This endeavor may contribute to social change by reinforcing the gap in practice evidenced in the literature relative to the use of the Istation digital formative assessment tool to influence student achievement. Investigating the use of a computer-based formative assessment program provides data analysis for teachers and administrators to apply when deciding whether such programs benefit the learner; helps teachers by guiding instruction; and creates a clear cycle through formative and summative assessment practices. I have provided a detailed description of the problem and purpose of the study and the research methodology and framework that guide the research questions. The end of the chapter includes the scope, limitations, assumptions, and delimitations of the study.

Background

Effective formative assessment strategies are key factors that, when used systematically, may double student growth (Polly et al., 2017). Luo, Lee, and Molina (2017) found a strong correlation between Istation individualized formative assessment and improved reading scores on the Standardized Test for the Assessment of Reading (STAR) in their mixed-methods study conducted with 98 third graders; they recommended further research with larger samples, especially if controlled for extraneous factors. Putman (2016) found a statistically significant relationship between the influence of the Istation reading program on kindergarten literacy achievement and indicated a need to conduct further research with other grade levels as these data included kindergarten only. Results from a study by Patarapichayatham, Fahle, and Roden (2014) found that Istation served as a predictor of how students perform on the State of Texas Assessments of Academic Readiness (STAAR) reading test; the authors suggested that additional research expand upon these findings. More recent studies found that Istation scores serve as a predictor of scores on other state summative tests and support the need for additional replication and a study that further isolates Istation's effect on student preparedness for summative exams (Campbell, Lambie, Sutter, Bickham, & Pulse, 2018; Patarapichayatham, 2018). Marin (2015) conducted a reading improvement study with Texas third graders' (n = 102) and suggested that Istation's effectiveness as appropriate formative assessment should be investigated in other grade levels and locations. To address the gap in practice indicated in the literature, I investigated the influence of the Istation reading program on Grades 3-5 students' preparedness and achievement on New

Mexico's ELA section of the PARCC summative assessment. This study may provide information to the stakeholders at the district about the function of the Istation formative assessment program to influence the mandated summative assessment. Data from this multigrade study in a different geographical location than previous studies may support Istation's influence on summative ELA assessments. This study was needed because investigating the problem addressed in this study may improve practice and narrow the gap in practice related to formative assessment in the education field.

Problem Statement

The research problem addressed in this dissertation was that the influence of the Istation reading formative assessment program on student achievement as measured by the ELA PARCC summative assessment is unknown. Students spend anywhere from 30 to 90 minutes a week interacting with the program, followed by a monthly assessment, yet the program's effects on reading proficiency for students in Grades 3-6 is unknown. Each year, the school's population at HES has a reading proficiency of about 30%. The district is using Istation to improve this figure, and it is crucial to find out its effects on reading proficiency and student growth. In appropriate data-based or data-driven decision making (DDDM), data about student learning guide teacher understanding of student progress and future instructional and curricular decisions (Bernhardt, 2016). Formative assessment requires a clear path from the evaluation of current skills to the level of skills that the student needs to have, with a significant emphasis on closing the gap in between (Sadler, 1989). Istation's designers posited that, with systematic use, their computer-based formative assessment reading program helps elementary students close the gap in

knowledge via individualized feedback on monthly assessments (Patarapichayatham, 2018). To further examine the influence of Istation's formative assessment on summative assessment and to address the gap in practice indicated in the literature, in this study I investigated the influence of the Istation reading program on Grades 3-6 students' achievement as measured by the scores on the ELA section of the PARCC summative assessment. Data from this multigrade study may provide support that the Istation program's individualized lessons help prepare students for summative ELA assessments and improve overall elementary reading skills. While formative assessment has been used in the education field for several years, as the education realm moves into automated data analysis, the influence of more recent digital assessment tools is significant to the curriculum, instruction, and assessment field (Bhagat & Spector, 2017). Recent literature included the potential of specific formative assessments, some of which are computer based, to improve certain aspects of reading student achievement and provide appropriate feedback to the student and teacher (see De Lisle, 2015; Karim, 2015; Petour, 2015). Previous researchers suggested further analyzing specific programs to assess the effectiveness of each one.

Purpose of the Study

The purpose of this quantitative study was to investigate the influence of the Istation reading formative assessment program on student achievement as measured by the ELA PARCC summative assessment. To address the gap in practice regarding the influence of the Istation formative assessment program on summative assessment, it was necessary to investigate how the students performed on the Istation Indicators of Progress

(ISIP) and whether their end of year ELA PARCC assessment scores deemed them proficient or not proficient. The results of the compared proficient or not-proficient groups may determine whether Istation's program offers formative data relative to influencing ELA performance. The state deems satisfactory student reading proficiency as PARCC scores at or above Level 4. The school administration provides Istation to enrich reading proficiency as part of the regular elementary curriculum; they may be able to use data from this study to guide program and function decisions about formative assessment processes. This research is also consistent with the Sadler (1989) framework that guides formative assessment practices and indicates that such individualized feedback may close the gap in knowledge between the student's current and desired levels of academic functioning. Specifically, I accessed and analyzed archived ISIP and ELA PARCC measures to investigate the effectiveness of the Istation program on reading achievement. The original data add to the existing research about the connection between formative and summative assessment.

Research Questions and Hypotheses

The purpose of this quantitative study was to investigate the influence of the Istation reading formative assessment program on student achievement as measured by the ELA PARCC summative assessment. Understanding Istation's effectiveness at preparing students to achieve satisfactory summative assessment scores is relative to addressing the gap in practice indicated in the literature as well as the theoretical framework in this study. The population for this study consisted of 281 students from a Title I school in third through sixth grade, although, as described in Chapter 4, the Grade

6 student data sets were excluded and removed prior to analysis. Each student in Grades 3-6 used Istation at least 25 minutes a week. Student data and progress were automatically collated into a student Istation report each month. Following each monthly formative assessment, the Istation's computer-based program informs students where they are regarding reading comprehension and understanding vocabulary and individualizes tasks to close the knowledge gap and move the student toward achieving grade-level skills. The state reviews and archives data from the Istation assessment at the beginning of the year (BOY), middle of the year (MOY), and end of the year (EOY). At the end of the school term, student achievement was cumulatively assessed with the ELA PARCC assessment.

Using Sadler's three-step process for effective formative assessment, the research questions guided the study to investigate whether there was a change in ISIP scores over time for students before, during, or after participation in the Istation reading program. The function of the Istation reading program aligns with Sadler's formative assessment framework by cycling students monthly through Sadler's steps: The first involves a monthly assessment (where the student is); Step 2 is goal setting (where the student needs to be); and Step 3 completes the feedback loop to evaluate goal attainment through continuous assessment (how to get students from where they are to where they need to be). A monthly assessment identifies current student level and creates a report to the teacher that indicates individual strengths and needs. Students complete assigned lessons until the next monthly assessment, where the program has filled the gaps in knowledge. Following the cycle theoretically better prepares students with individualized feedback

for summative reading assessments. This framework informed the research questions and analysis for this study by providing a lens through which to analyze student outcomes from Istation data. The following research questions guided this study:

RQ1: Are there statistically significant differences in ISIP scores for Grade 3 before (BOY), during (MOY), and after (EOY) participation in the Istation reading program?

$H_0$1: There are no statistically significant differences in ISIP scores for Grade 3 before (BOY), during (MOY), and after (EOY) participation in the Istation reading program.

H_a1: There are statistically significant differences in ISIP scores for Grade 3 before (BOY), during (MOY), and after (EOY) participation in the Istation reading program.

RQ2: Are there statistically significant differences in ISIP scores for Grade 4 before (BOY), during (MOY), and after (EOY) participation in the Istation reading program?

$H_0$2: There are no statistically significant differences in ISIP scores for Grade 4 before (BOY), during (MOY), and after (EOY) participation in the Istation reading program.

H_a2: There are statistically significant differences in ISIP scores for Grade 4 before (BOY), during (MOY), and after (EOY) participation in the Istation reading program.

RQ3: Are there statistically significant differences in ISIP scores for Grade 5 before (BOY), during (MOY), and after (EOY) participation in the Istation reading program?

$H_0$3: There are no statistically significant differences in ISIP scores for Grade 5 before (BOY), during (MOY), and after (EOY) participation in the Istation reading program.

H_a3: There are statistically significant differences in ISIP scores for Grade 5 before (BOY), during (MOY), and after (EOY) participation in the Istation reading program.

RQ4: To what extent do Istation formative assessment scores (MOY) predict literacy performance as measured by the ELA PARCC summative assessment for Grade 3?

$H_0$4: The Istation formative assessment does not predict literacy performance as measured by the ELA PARCC summative assessment for Grade 3.

H_a4: The Istation formative assessment does predict literacy performance as measured by the ELA PARCC summative assessment for Grade 3.

RQ5: To what extent do Istation formative assessment scores (MOY) predict literacy performance as measured by the ELA PARCC summative assessment for Grade 4?

$H_0$5: The Istation formative assessment does not predict literacy performance as measured by the ELA PARCC summative assessment for Grade 4.

H_a5: The Istation formative assessment does predict literacy performance as measured by the ELA PARCC summative assessment for Grade 4.

RQ6: To what extent do Istation formative assessment scores (MOY) predict literacy performance as measured by the ELA PARCC summative assessment for Grade 5?

H_06: The Istation formative assessment does not predict literacy performance as measured by the ELA PARCC summative assessment for Grade 5.

H_a6: The Istation formative assessment does predict literacy performance as measured by the ELA PARCC summative assessment for Grade 5.

I analyzed the quantitative research questions to find out whether the Istation scores changed over time and whether they influenced each student's summative assessment (ELA PARCC). The dependent variable for the first three research questions in this study was the Istation formative assessment program, while the independent variables were the ISIP scores from the three points in time throughout the school year (BOY, MOY, EOY). The dependent variable for the last three research questions in this study was the ELA PARCC and the independent variable that was investigated as a predictor was the MOY ISIP score from the Istation reading program. Determining whether these data sets influenced student preparedness for summative assessment may confirm the effectiveness of Istation's formative assessment process, thereby addressing a gap in practice identified in the literature. Providing data to support Sadler's (1989) framework is critical to make instructional decisions that close the gap between current academic knowledge and the ideal academic goals for a classroom of students.

Theoretical Foundation

Sadler's (1989) 3-Step process for effective formative assessment was the theoretical framework for this study. Sadler (1989) indicated that designing instruction in accordance with the formative assessment process increases student achievement on summative assessments. Sadler's research built upon the study by Ramaprasad (1983) which discussed the definition of feedback as the input, existence of data, and the gap between the actual and referential level of skills. Sadler posited the need to establish where the learners are in their learning (Step 1), where they are going (Step 2), and what needs to be done to get them there (Step 3). Sadler's 3-step process frames the formative assessment cycle, designed to ultimately improve student achievement (Black & Wiliam, 1998). Ascertaining that student learning is the output of Sadler's 3-Step process is an important factor.

Sadler's 3-Step formative assessment process was an appropriate framework for this study because it is unknown whether the Istation formative assessment program effectively prepares students for the required summative assessments. Determining the alignment between Istation reading scores and student summative scores may provide quantitative data indicating that the framework is effectively guiding instruction and learning in the classroom. The Istation reading program was designed to follow Sadler's formative assessment framework by cycling students monthly through Sadler's steps: a monthly assessment, goal setting, and completing the feedback loop to evaluate goal attainment through continuous assessment (Sadler, 1989). A monthly assessment evaluates existing student skills and level (Step 1), and a printed goals-report includes

identified skills needed to meet grade-level content standards (Steps 2 and 3). Following each assessment, the Istation program software automatically assigns individual lessons to help predict what the student needs to fill the gaps in knowledge. Determining whether the data sets from Step 3 influence student preparedness for summative assessment may confirm the effectiveness of Istation's formative assessment process, thereby addressing a gap identified in the literature. The formative assessment process and the previous works that follow the framework proposed by Sadler (1989) are detailed in Chapter 2. Providing data to support Sadler's framework is critical to make instructional decisions that close the gap between current academic functioning and the ideal academic goals for a classroom of students.

Nature of the Study

The focus of this study was to determine the extent of the influence of the Istation reading individualized, computer-based, formative assessment program on student achievement as measured by the ELA PARCC summative assessment. First, I chose a quantitative design using the one-way repeated measures analysis of variance (ANOVA) because the study involved measuring each student's data over the three time points using the same measurement at each time point (Laerd Statistics, 2019). Secondly, I chose a quantitative design using the linear regression because the study involved investigating the relationship between the Istation formative assessment program and the summative ELA PARCC summative assessment. The Istation score informs the students where they are regarding reading comprehension, spelling, text fluency, and understanding vocabulary, and individualizes tasks to close the gap and move the student toward

achieving grade level skills. As described in Chapter 4, the analyzed data set consisted of scores from 175 students from a Title I school in Grades 3 through 5. For the one-way repeated measures ANOVA, the 2017–2018 ISIP score received on the Istation's computer-based program was the dependent variable. The independent variable was the ISIP score from each of the three points in time that students took the Istation formative assessment in 2017–2018. For the linear regression, the dependent variable was the ELA PARCC score and the independent variable was the Istation reading formative assessment MOY ISIP score, assessed to see its predictive ability on the ELA PARCC summative test.

Definitions

Formative assessment: Refers to a wide variety of methods that teachers use to conduct in-process evaluations of student comprehension, learning needs, and academic progress during a lesson, unit, or course (Great Schools Partnership, 2014). Sadler (1989) stated that it is "how judgments about the quality of student responses (performances, pieces, or works) can be used to shape and improve the student's competence by short-circuiting the randomness and inefficiency of trial-and-error learning" (p. 120).

High-Stakes Assessment: Assessment which is summative in nature and is used to rate the performance of a student, teacher, leader, and/or school. This can include state end of year assessments (McDowell, Smailes, Sambell, Sambell, & Wakelin, 2008)

Istation's Indicators of Progress (ISIP): ISIP is an "Internet- and Web-delivered computer-adaptive testing system that provides continuous progress monitoring

assessments in the critical domains of reading in prekindergarten through eighth grade" (Patarapichayatham et al., 2014, p. 3).

Partnership for Assessment of Readiness for College and Careers Assessment (PARCC): The PARCC is an assessment that a group of states worked together to develop. The PARCC measures whether students are on track to be successful in college and careers. The statewide standardized examination is used to test students. The PARCC is based on the federal Common Core standards, and it tests students on the curriculum goals for each respective grade, including the areas of ELA and mathematics in Grades 3–11 (New Jersey Department of Education, 2015).

Schoolwide Title I: As part of the Every Student Succeeds Act (ESSA), a Title I school receives financial assistance due to the high percentage of children served that come from low-income families (New Jersey Department of Education, 2015).

Summative Assessment: A summative assessment is administered to students at the end of an instruction cycle to certify students or curriculum (Black & Wiliam, 2009).

Assumptions

To produce reliable results for this study, it is necessary to recognize and share assumptions and their potential impact. The first assumption was that the principal of the school kept the PARCC data and Istation monthly ISIP scores in a password-protected computer. With the Istation program being relatively new to the school (two full years of adoption), an assumption was that the teachers administered the test to each student every month of the school year and that students were given the opportunity to meet their suggested minutes on the Istation program each month. A last assumption is that teachers

and administrators ethically proctored the summative ELA PARCC assessment in various locations throughout the month-long testing period.

Scope and Delimitations

This study was limited to elementary and intermediate school reading. Therefore, the results are limited to one content area in multiple grades. Additionally, the scope of this study was concentrated on a student population in Grades 3 through 6 in a Title 1 school district comprising 281 students in a rural southwestern town. (However, due to incomplete data sets, I excluded all Grade 6 data sets from the data analysis, adjusting the population from 281 to 201 and from Grades 3-6 to 3-5, as discussed in further detail in Chapter 4.) Since I work at the school where the study took place, the location and number of students in the sample was one of convenience. The study was limited to the indicated grade-levels and the reading content area, using archived data. The specific focus was chosen because of the limited amount of research regarding the Istation formative assessment's effects on summative assessment for those grade levels. These results, provided in Chapter 4, may generalize to other Title I schools with students in Grades 3 through 6 in the same region or in similar regions.

Limitations

One possible limitation could be that I was working at the school from which these data were retrieved. However, since these data are archival, there was no conflict of interest. These data that I analyzed were from one Title I school in one school district in the Southwest region of the United States. Therefore, the findings may not be generalizable to students in these same grades in other regions. Also, through this study I

provided information about the influence of Istation's scores and not necessarily the quality of its use. Finally, though I found some articles and dissertations related to the Istation program (the independent variable in the study), none of these studies listed such measures. To gather this information, I contacted the Istation designers but have not received a reply. Therefore, a final limitation was the lack of published validity and reliability measures of the Istation program. As I accessed data that were already deidentified, there is not a concern for confounding variables such as age, gender, or grade level.

Significance

This study was important to address the gap in practice relative to the Istation individualized computer-based formative assessment's ability to influence the PARCC summative assessment and improve student achievement. The study may benefit the education field as it may provide data to indicate Istation's ability to influence student performance in reading on the PARCC summative assessment. The study may contribute to addressing the gap in practice related to assessment by offering an understanding of the students' Istation program's formative assessment scores' ability to prepare students for the end of year summative assessment, the ELA PARCC. HES has students who take the PARCC assessment and are using the Istation program without evidence of its ability to prepare students to achieve proficiency on the PARCC. Teachers may benefit from this information as it may provide formative data for students needing support and offer areas for remediation or extension. Teachers may gain a valid, reliable tool to understand strengths, areas for growth, and the ability to personalize learning. Black and Wiliam

(2009) found that it is crucial for the formative assessment to include interactive feedback. Istation's designer aimed to accomplish this task in its reading formative assessment program. The study may help to determine whether there is a significant relationship between achievement of the students using the Istation ISIP reading formative assessment and interactive feedback versus teacher-led reading formative assessment. Results may provide information to the education field about the effectiveness of the program and make informed decisions regarding its continued use in school districts.

The study findings may indicate efforts to close the gap in practice at the local school by indicating the function of the Istation formative assessment program relative to student achievement on the PARCC. Furthermore, because the Istation tool was mandated by the state to be used in Grades K-2, the results may inform stakeholders about the necessity of the program's use as an effective formative assessment tool for upper elementary students. The effects of closing this gap in practice involve how teachers and students use formative assessment data to guide instruction. The study contributes to the literature about the program's use and supports the practical application of digital formative assessment use in the education field. This endeavor may contribute to social change by reinforcing the gap in practice relative to the use of a digital formative assessment tool to influence student achievement. Investigating the use of a computer-based formative assessment program provides data analysis for teachers and administrators to apply when deciding whether such programs benefit the learner, help

teachers by guiding instruction, and create a clear cycle through formative and summative assessment practices.

Summary

I investigated the influence of the Istation reading program on student preparedness and achievement as measured by the ELA PARCC summative assessment. Using Sadler's three-step process, I determined to what extent the Istation formative assessment program scores influence the summative assessment for students attending HES. The information gained may contribute to the gap in practice evidence in the literature by linking formative to summative assessment. In the next chapter, I explore current research to provide a background for the study in the form of a literature review.

Chapter 2: Literature Review

The research problem addressed in this dissertation was that the influence of the Istation reading formative assessment program on student achievement as measured by the ELA PARCC summative assessment is unknown. The purpose of this quantitative study was to investigate the influence of the Istation reading formative assessment program on student achievement as measured by the ELA PARCC summative assessment. In a mixed methods study with 98 third-grade children, Luo et al. (2017) found a strong correlation between Istation individualized formative assessment and improved reading scores on the STAR; they recommended further research with larger samples, especially if controlled for extraneous factors. Putman (2016) found a statistically significant relationship between the effect of the Istation reading program and kindergarten literacy achievement and indicated a need to conduct additional studies with other grade levels and locations. Results from a study by Patarapichayatham et al. (2014) found that Istation served as a predictor of how students perform on the STAAR reading test; the authors suggested that additional research expand upon these findings.

More recent studies found that Istation scores serve as a predictor of scores on other state summative tests and support the need for additional replication and a study that further isolates Istation's effect on student preparedness for summative exams (Campbell et al., 2018; Patarapichayatham, 2018). Marin (2015) conducted a reading improvement study with Texas third graders' (n = 102) and suggested that Istation's effectiveness as appropriate formative assessment should be investigated in other grade levels and locations. In this chapter, I include strategies for the literature search, the

theoretical foundation of the study, and a comprehensive literature review related to key concepts and variables. The literature review indicates the need for additional research regarding the Istation program's formative assessment's relationship to summative assessment.

Literature Search Strategy

The literature presented in this review focuses on formative assessment practices, recent changes in the field including the use of computer-based and individualized formative assessment for reading, and research on the link between formative and summative assessments. Researching the background of formative assessment and how it has evolved since its inception helped structure the search and funnel down to the effects of specifically the Istation program. In this next section, I include a list of library databases and search engines used, key search terms and combinations of search terms, and the scope of the literature review in terms of years and types of literature searched.

Databases and Search Engines Used for Review

The databases I researched for this literature review include Academic Search Complete, ERIC, Education Source, and ProQuest. I also found some dissertations related to the program's influence on ELA achievement for kindergarten through Grade 2 on the Istation website. The main keyword phrases used to search were *Istation, formative assessment effectiveness, summative assessment, elementary or primary education,* and *student achievement, Grades K-8, student assessment, the PARCC exam, computer-based assessment, computer-based formative assessment, predictors of achievement, game-based learning, digital learning,* and the works of *Black and Wiliam,*

Ramaprasad, and *Sadler*. However, since the Istation program was only founded in 1998, I also researched how previous formative assessment programs influenced summative assessment in the education field. Key phrases included a combination of two or more of the following: *formative assessment and summative assessment*, *reading or literacy,* and *computer-based formative assessment,* and *student achievement*.

Scope of Review

Other than the formative assessment seminal works of Ramaprasad (1983), Sadler (1989), and Black and Wiliam (1998), I limited the search to current, peer-reviewed studies from the previous 5 years, 2014–2019. Among these studies are three comprehensive literature reviews of several formative assessment studies and four robust reviews of computer-based formative assessment. The Black and Wiliam formative assessment background paired with the current studies provide the theoretical foundation and current need for additional research.

Theoretical Foundation

The work of Sadler (1989) and his 3-Step cycle frames the study. Educators and policy makers view formative assessment as a primary approach to educational reform (Herman, Osmundson, Dai, Ringstaff, & Timms, 2015; Petour, 2015). Educators are seeking ways to understand what skills the students have so they can, ultimately, perform better on high-stakes assessments. With the focus on formative assessment practices, Sadler's 3-Step cycle aligns with the way the Istation's designers posit their program works to improve student achievement.

Origin and Hypotheses of Sadler's Theory

Sadler (1989) indicated that designing instruction in accordance with the formative assessment process increases student achievement on summative assessments. The author built upon the study by Ramaprasad (1983), who defined *feedback* as the input, existence of data, and the gap between the actual and referential level of skills. Sadler posited the need to establish where the learners are in their learning (Step 1), where they are going (Step 2), and what needs to be done to get them there (Step 3). Sadler's 3-step process frames the formative assessment cycle with the hypothesis that it will ultimately improve student achievement (Black & Wiliam, 1998). Sadler shared that effective formative assessment used with this cycle can help students fill the gap in knowledge. Ascertaining that student learning is the output of Sadler's 3-Step process is an important factor.

Rationale

The framework theorized by Sadler (1989) is appropriate because his 3-Step formative assessment process outlines the cycle intended to close the gap in knowledge. Since Istation's designers claim to be able to close this gap in knowledge for each student via personalized and individualized feedback, the program's use aligns with the three steps the framework outlines: from where the student is, to where they should be, and how to get them there. Determining the alignment between Istation reading scores and student summative scores would provide quantitative data indicating that the framework is effectively guiding instruction and learning in the classroom. The Istation reading program was designed to follow Sadler's formative assessment framework by cycling

students monthly through Sadler's Steps: a monthly assessment, goal setting, and completing the feedback loop to evaluate goal attainment through continuous assessment. A monthly assessment evaluates existing student skills and level (Step 1), and a printed goals report includes identified skills needed to meet grade-level content standards (Steps 2 and 3). The Istation program software assigns individual lessons to help predict what the student needs to fill the gaps in knowledge. Following the cycle theoretically better prepares students with individualized feedback for summative reading assessments.

Application From Previous Literature

Previous authors of formative assessment studies included Sadler's 3-Step process. In their article about assessment for learning, Calfee, Wilson, Flannery, and Kapinus (2014) shared the importance of student involvement and feedback in the learning process (Sadler, 1989). Das et al. (2017) conducted research that involved medical students using formative assessments. The authors found that feedback is important to help the students fill their learning gaps (Das et al., 2017). Dixson and Worrell (2016) used Sadler's specific language to describe how teachers and students can use formative assessment to improve student achievement on summative assessment. The study in hand is like these studies because I used Sadler's framework to investigate the relationship between student performance on the Istation program formative assessment and the ELA PARCC assessment.

Relevance to Research Questions

This framework informs the research questions and analysis for this study by providing a lens through which to analyze student outcomes from Istation data.

Determining whether the data sets from Steps 1-3 influence student preparedness for summative assessment may confirm the effectiveness of Istation's formative assessment process, thereby addressing a gap identified in the literature. Finding out whether there is a change in ISIP score for students who did or did not receive a proficient score on the PARCC builds upon the existing theory of Sadler's 3-Step formative assessment process aimed at closing the gap in knowledge. Providing data to support Sadler's (1989) framework is needed to make instructional decisions that close the gap between current academic knowledge and the ideal academic goals for a classroom of students.

Literature Review Related to Key Concepts and Variable

Formative assessment is widely used to assess student learning and guide instruction. In this section, I include information related to formative assessment in general and recent research in the field related to the components of this study: the link to summative assessment and student achievement in various areas including a comprehensive group of studies specifically about reading, a review of some computer-based assessments and literacy, and finally the current research on the Istation program related to all the above information. The research may indicate the need for the study of the specific programs and their ability to influence student achievement.

Formative Assessment and Feedback

Formative assessment provides feedback to students to help them learn effectively, and it can come in many forms (Black & Wiliam, 2009). Formative assessments also inform students and their teachers about what the learning goal is, where the students are in relation to that learning goal, and what can be done to improve

subsequent performance (Black & Wiliam, 2009; Dixson & Worrell, 2016; Sadler, 1989). Recent studies indicated the potential of formative assessment to provide such feedback about student performance (De Lisle, 2015; Herman et al., 2015; Karim, 2015). A university study by Owen (2016) revealed that it is important to offer students multiple, low-stakes, active learning opportunities. The monthly Istation program can be viewed as low stakes because it is not part of the student end-of-year portfolio of assessments. However, formative assessment is not always congruent with teachers' pedagogical beliefs and practices. Researchers who sought teachers' perspective and evaluation of assessment practices found that even though educators and policy makers historically viewed formative assessment as a main approach to educational reform, teachers need to have ownership of assessment practices for student success and a clear system for feedback (De Lisle, 2015; Herman et al., 2015; Petour, 2015; Sach, 2015). These studies indicate the need for additional research in specific practices.

Formative assessment practices and teachers' understanding of executing such practices are not always alike. A study completed by Cotton (2017) found that students rated teachers higher than teachers rated themselves on formative assessment practices. Curry, Mwavita, Holter, and Harris (2016) found that formative data use facilitates teacher motivation but, to be effective, must be completed under the proper conditions, including district-wide support. Clinchot et al. (2017) stated that prescribed assessment can miss opportunities to understand students' reasoning. The authors suggested the need for a four-step, responsive formative assessment where information is elicited, noticed, interpreted, and acted upon to elicit students' way of thinking (Clinchot et al., 2017).

Srivastava, Waghmare, and Mishra (2018) reported positive effects when investigating the use of formative assessment with first-year medical students. Using clearly defined formative assessment classroom techniques (FACTs), the effect size for student achievement of the treatment group was 1.12 (Srivastava et al., 2018). The authors found that the FACTs are useful tools to assess understanding, tailor instructional modifications, and facilitate feedback (Srivastava et al., 2018). Drawing on the recent studies and the seminal theories of Black and Wiliam (1998) and Sadler (1989; 1998), the researchers indicated the need to investigate specific formative assessment strategies in today's classrooms. This research informs the current study because Istation is a specific formative assessment program with little data to indicate its effectiveness in the classroom setting for students in Grades 3 through 6.

The Synergy of Formative and Summative Assessment

Formative and summative assessment should work in tandem to achieve positive student outcomes. Dixson and Worrell (2016) stated that formative assessment is "a whole set of tools to provide feedback to help students learn more effectively" (p. 154), and summative assessments are high stakes because they seek to gather a final snapshot of the student's achievement that school year. However, in the field of education research, formative assessment has been coined the *good* assessment while summative assessment has been viewed as *bad* (Lau, 2016). Buelin, Ernst, Clark, Kelly, and DeLuca (2019) used a doctor's visit analogy to describe the dichotomy. Formative assessment can be viewed as a check-up at the doctor to evaluate symptoms, while the summative assessment is more like an autopsy (Buelin et al., 2019). Upon reviewing the literature,

Lau (2016) cautioned policy makers and stakeholders against this dichotomy. The author shared the need to consider the fundamental idea of the two assessments harmoniously working to promote student achievement (Lau, 2016). The school where I conducted this study purchased an Istation program license for Grades 3 through 6 in hopes that the program would improve achievement and overall reading skills for the students. Understanding the link between the Istation formative assessment and future summative assessments can help teachers understand how to use them in daily practice.

The Effectiveness of Formative Assessment

Over several years, many authors published studies about the effects of various formative assessment programs leading to significant, and sometimes conflicting, information. Some of these researchers investigated the difference in student achievement in students who received formative assessment strategies and those who did not. Between 2014 and 2019, researchers wrote three articles that included meta-analyses of the previous literature. Specifically, Klute, Apthorp, Harlacher, and Reale (2017) conducted a comprehensive review of 19 studies of elementary schools and their students based on the formative assessment framework by Black and Wiliam. Several findings indicated that students who participated in formative assessment performed better on measures of academic achievement at the elementary level. Alternately, a review of over 160,000 studies conducted by Apthorp, Klute, Petrites, Harlacher and Real (2016) revealed the need for standards to assess whether formative assessment is effective on student achievement. Though the studies reviewed had evidence, assigning each study a score of having met, partially met, or not met the Procedures and Standards Handbook indicated

that many did not necessarily strongly support formative assessment procedures (Apthorp et al., 2016). Xu and Brown (2016) looked at several articles regarding teacher assessment literacy and found that tensions and other political issues with stakeholders impede the progress of teachers becoming versed in assessment practices. The authors suggested viewing assessment literacy as a continuum to allow teachers to evolve in their assessment practices as opposed to labeling them *literate* or *illiterate* (Xu & Brown, 2016). These reviews of formative assessment provide a base to work from when reviewing specific formative assessment programs.

Specific Formative Assessment and Mathematics

Some authors of individual studies showed potential for specific formative assessment programs and practices to prepare students for summative testing in mathematics (Andersson & Palm, 2017; Das et al., 2017; Polly et al., 2017; Ravenel, Lambeth, & Spires, 2014). Andersson and Palm (2017) found that following professional development on using formative assessment to teach math, Grade 3 students' math achievement improved on the high-stakes test. Similarly, Polly et al. (2017) found that the treatment group who received formative assessment on number sense tasks demonstrated growth on the summative assessment. That growth was more substantial the more times the formative assessment was used to collect data (Polly et al., 2017). Das et al. (2017) found that, overall, medical students found their use of formative assessment practices helpful to minimize their learning gap. However, the researchers did not quantitatively analyze whether their scores improved on summative scores. In Chapter 3,

I outline a way to quantitatively assess the connection of student performance on reading summative assessment.

Specific Formative Assessment and Other Subjects

Additional studies include authors finding improvement in achievement following formative assessment in other content areas (Aydin & Ürün, 2016; Huang, 2016; Ozan & Kincal, 2018). Aydin and Ürün (2016) found a significant difference between pre- and posttest scores with the use of formative assessment in a science unit about space in comparison to a control group where teachers did not administer formative assessment strategies. A theater arts study by Huang (2016) saw a positive effect of formative tests on student achievement and rote memorization of content. The results from a Grade 5 social studies course with 45 students conducted in Erzurum, Turkey, indicated that students in the formative testing experimental group had both a significantly higher academic achievement rate and better attitudes about the content (Ozan & Kincal, 2018). These studies share the theme that formative assessment practices can be a positive addition to various education environments and, in some cases, improve student achievement. This research supports the idea that assessing individual programs can benefit the stakeholders using them on a regular basis, which is what I investigated in my study.

Alternately, some studies revealed that formative assessment practices do not always prepare students for the summative assessment of the specific subject assessed (Bulunuz, Bulunuz, Karagöz, & Tavsanli, 2016; Grosas, Raju, Schuett, Chuck, & Millar, 2016; Pinger, Rakoczy, Besser, & Klieme, 2018). Pinger et al. (2018) found that students

who used formative assessment tools did not always show growth in math and that students may require more detailed material and guidelines. Rakoczy et al. (2019) found feedback to be more useful in a mathematics course for participants in the formative assessment condition versus the control, but the learning progress of the groups did not differ. Grosas et al. (2016) investigated formative assessment given to postsecondary students taking an immunology course. The researchers found that despite overall excitement and motivation regarding the formative tests, the students' scores on the summative exam were disappointing (Grosas et al., 2016). Bulunuz et al. (2016) conducted a comparison of science formative assessment and summative assessment and found that formative multiple-choice questions did not help students' achievement on the summative science exam. Regarding formative assessments for preservice teachers, Matthews and Noyes (2016) argued that it is more important to investigate where the students are rather than their particular grade on a formative test. In these cases, the researchers did not reveal a clear relationship between formative assessment and student achievement.

Other Formative Assessments and Literacy

Specific to the literacy or reading content area, some studies show academic achievement improvement in the key components of literacy (Barefoot, 2017; Boumediene & Hamzaoui-Elachachi, 2017; Bulat et al., 2017). Bulat et al. (2017) reported that formative assessment is widely used as part of the response to intervention model in many low- and middle-income countries as part of their feedback cycle. The author noted that formative assessment in the classroom substantially improves literacy

outcomes (Bulat et al., 2017). In their study of text comprehension in Algeria, Boumediene and Hamzaoui-Elachachi (2017) found that formative assessment processes showed significant progress in grammar, textual, functional, and sociolinguistic language. Barefoot (2017) and Bennett, Gardner, Cartledge, Ramnath, and Council (2017) found significant improvement in their participants when using formative assessment practices. Specifically, students taking a library course improved their motivation for conducting research versus the control group (Barefoot, 2017). Bennett et al. (2017) saw a positive effect of formative testing on fluency and comprehension skills. The researchers also found an improvement on the Aimsweb assessment following the treatment (Bennett et al., 2017). These researchers saw positive effects when using formative assessment with their participants, as the Istation program designers claim to accomplish with their formative assessment program.

Formative Assessment and Diverse Populations

Formative assessments show an improvement in skills for students of color (Council, Cartledge, Green, Barber, & Gardner, 2016; Li, 2016). In a reading intervention study for Grade 2 African American students at risk in an urban area, Council et al. (2016) conducted a similar study with three African American students in an urban school and determined that the students' reading and behavioral outcomes improved following implementation of the formative assessment intervention. Li (2016) found that reading achievement improved for black students more than white when using formative assessment and found that these results showed potential for formative assessment to reduce achievement gaps between students of different races (Li, 2016). Two-thirds of the

population at HES are Native American, Latino, or other mixed races. Formative assessment practices could improve learning for the diverse population at the school.

Formative assessment can improve reading comprehension and decoding in various grade levels (Dupont, 2018; Gustafson, Nordstrom, Andersson, Falth, & Ingvar, 2019; Simmons et al., 2015). Dupont (2018) conducted a reading study related to formative testing via workstations with Grades 6 and 8 students in France. These workstations improved reading comprehension and were believed to help close the gap in learning. Similarly, Simmons et al. (2015) saw positive effects of a supplemental formative reading intervention program in kindergarten. Students whose scores on the tests were above 90% accelerated in their studies while others who had lower than 70% achievement repeated lessons until they were able to progress (Simmons et al., 2015). The researchers found a steady progression of the curriculum for students taking these tests. Finally, the study of a program called LegiLexi in Sweden by Gustafson et al. (2019) saw good outcomes for formative assessment: an increase in student ability to decode and comprehend texts. These studies indicate the potential of formative assessment practices to improve student learning and bring students to where they need to be in their learning path, just as Sadler's framework suggests for a working formative assessment cycle.

Istation's designers hold that the program individualizes lessons to improve reading student achievement. There are two studies that showed little to no effect of formative testing on reading student achievement (Faber & Visscher, 2018; Saito & Inoi, 2017). The software includes a spelling component in its formative assessment and

individualized instruction lessons. Formative assessment practices did not affect spelling achievement with a population of Grade 3 students (Faber & Visscher, 2018). The study indicated the need for future research in the areas of student feedback and differentiated instruction. Saito and Inoi (2017) conducted research with junior and senior high school students learning English as a foreign language and found that teachers used formative assessment to submit as grades while forgetting about the feedback potential for students. However, the study indicated a concern in that the teachers used formative assessment to varying degrees and shared teacher training, intent, and purposes of the test as implications (Saito & Inoi, 2017). Based on the research of recent literacy-based studies, it is necessary to conduct further research in formative testing and its effects on reading student achievement. Istation is one of the more popular reading formative assessment programs with little research of effectiveness on student achievement.

Computer-Based Formative Assessment

In a world with technology at everyone's fingertips, formative assessment via computer-based programs continues to grow in the education field. Timmis, Broadfoot, Sutherland, and Oldfield (2015) conducted a research review that discusses the possibilities of technology-enhanced assessment (TEA) in classrooms. The authors found that with the use of TEA, formative and summative assessment can become "more relevant to learners," yet stressed the need to amend policies to avoid "patchy incremental change" (Timmis et al., 2015, p. 468). In a review of nine review papers and eight empirical studies, Shute and Rahimi (2017) found that complex competencies in various content areas can be measured by more recent computer-based assessment for learning.

Similarly, Belo, McKenney, Voogt, and Bradley (2016) conducted a review of the literature surrounding computer-based formative assessment programs and their effects on early literacy. The authors found aspects of specific applications to be helpful, including phonics and vocabulary programs, electronic storybooks, tutorials, and narrative educational television shows (Belo et al., 2016). Specifically related to teacher inquiry, Luckin, Clark, Avramides, Hunter, and Oliver (2016) conducted a literature review and share the need to enable teachers to use technology effectively toward student growth. Studies in Sweden, China, and Australia explored technology related to formative feedback in the elementary setting and found the potential in the use of digital learning (Bhagat & Spector, 2017; Cloonan, Hutchison, & Paatsch, 2016; Genlott & Grönlund, 2016). These authors shared the suggestion of additional research about specific programs and their ability to properly assess learning.

Computer-based formative assessments often include interactive activities. The way the Istation program tests students is via game-based assessments where students complete missions within the realm of comprehension, spelling, vocabulary, and text fluency. In a study of nursing students who used a game show quiz-style approach to formatively assess learning, Aljezawi and Albashtawy (2015) found no difference in student achievement in pre- and posttest scores with formative assessment versus lectures (Aljezawi & Albashtawy, 2015). However, results indicated that the students who took the quiz felt it was a more satisfying instructional method, and their immediate feedback indicated an increase in information retention (Aljezawi & Albashtawy, 2015). The Istation program's design also offers immediate feedback in a game-style manner.

Learning outcomes are important when considering game-based activities to assess learning. When reviewing components of digital game-based learning activities, All, Nunez Castellar, and Van Looy (2015) found that the tasks must be related to learning outcomes and real-world situations. The authors stated the need for developers to be aware of specific learning targets when designing such activities (All et al., 2015). The Istation program includes a standards-based component that links each task to a similar CCSS learning standard to theoretically help students by filling in the gap in knowledge as they go through the prescribed lessons.

Recent articles indicate that technology can be used to support formative assessment of learning standards (Gallagher, 2016; Martin, Polly, Chuang, Lambert, & Pugalee, 2016; Zlatovic, Balaban, & Kermek, 2015). Martin et al. (2016) found a statistically significant increase in teacher practices to be more student centered. Also, the researchers found that technology has the potential to impact instructional decisions and provide an easier way to analyze data (Martin et al., 2016). Gallagher (2016) reported similar positive effects of technology in helping teachers improve the efficiency of data collection and personalize learning. Zlatovic et al. (2015) found that students who received postassessment feedback online could steer their future learning strategies. These authors shared the use of feedback like Sadler's Step 3. Istation's designers also claim that the program personalizes learning and provides feedback.

Teachers and students can benefit from immediate feedback and individualized data (Shirley & Irving, 2015; Spector et al., 2016). In the study by Shirley and Irving (2015), high school science teachers used online assessments to generate robust evidence

of individual student learning that they then could use to make instructional decisions. Spector et al. (2016) found that technology helps improve student motivation and engagement and can be adapted and differentiated to target student deficiencies. Similarly, students receive personalized feedback from the Istation program that then is shared with the teacher about progress. These articles provide support for the idea that computer-based programs like Istation can offer individualized, efficient data to the student and teacher to provide personalized feedback and drive future instruction in various content areas.

Computer-Based Formative Assessment and Literacy

Based on recent literature, digital formative assessment can improve the specific and important content area of student reading and literacy. Buysse et al. (2016) reviewed two studies about the formative assessment Recognition & Response. After reviewing the teachers' ability to differentiate instruction and individualize student feedback with fidelity, authors found that the Recognition & Response small-group lessons showed larger gains than the control group (Buysse et al., 2016). This research supports the idea that personalized feedback can help students learn.

Similarly, in a meta-analysis of computer-based assessments' affect on student achievement in writing skills for students in Grades 1 through 8, Graham, Hebert, and Harris (2015) found that formative assessment statistically enhanced writing quality. The computer program offered feedback directly to the students, and the authors concluded that formative assessment can be beneficial to improving student achievement (Graham

et al., 2015). The Istation designers claim to complete the same individualized feedback in the formative assessment program to complete the loop in Sadler's Step 3 (1989).

Comprehension is a key component to reading and is evaluated via Istation's formative assessment program. Hooley and Thorpe (2017) found that, following formative assessment via an online program, high school students improved content comprehension and motivation to read informational texts. Similarly, Gustafson, et al. (2019) found that, with full access to the online formative assessment program LegiLexi, Grade 3 students improved their comprehension and decoding. The treatment group used only the formative assessment tool, and the students improved only in comprehension (Gustafson et al., 2019). These researchers indicated the ability of formative assessments to improve comprehension for young learners and high schoolers. Little recent research exists with reading comprehension via formative assessment for other grades.

Text fluency is another integral component of reading. Some studies found an increase in oral reading fluency when using computer-based formative testing (Bennett et al., 2017; Keyes, Cartledge, Gibson, & Robinson-Ervin, 2016). Bennett et al. (2017) and Keyes et al. (2016) evaluated Grade 2 students and found that computer-based formative testing improved oral reading and fluency. Specifically, five out of six at-risk Grade 2 students improved their oral reading fluency (Keyes et al., 2016). Bennett et al. saw a positive effect on fluency and comprehension with Grade 2 African American students. Though they were small studies, the authors indicated the ability of some computer-based programs to help improve fluency for young learners.

Conversely, some studies show that computer-based measures do not necessarily improve reading skills for learners (Clemens et al., 2015; Faber & Visscher, 2018); Fenty, Mulcahy, & Washburn, 2015). Researchers found an improvement in fluency, letter naming, and letter sound identification via paper-based measures versus computer-assisted learning (Clemens et al., 2015). Faber and Visscher (2018) assessed Grade 3 students on spelling achievement via online formative assessment and found that teachers need to look at specific student feedback. The researchers saw no improvement in spelling achievement via online testing measures (Faber & Visscher, 2018). Fenty et al. (2015) found no difference between teacher-led and computer-based fluency instruction with 50 Grade 3 students whose fluency development was delayed. These researchers suggested that studies need to look at specific programs when assessing the ability of formative assessment to improve literacy. These studies indicate the need to further evaluate formative assessments via computer-based programs.

Learners growing up in the 21st century require tools that match their learning. Cloonan et al. (2016) conducted a study in Australia and shared that need for educators to embrace digital learning tools to help teach literacy to our 21st-century learners. The authors reported seven affordances of digital learning to teach literacy, including active learning making, differentiated instruction, metacognition, collaborative intelligence, recursive feedback, ubiquitous learning, and multimodal knowledge representations (Cloonan et al., 2016). The authors showed these affordances in relation to the characteristics of effective formative assessment outlined by Black and Wiliam (1998). The authors stated that digital assessment offers greater, more diverse opportunities than

the standard print-based classrooms (Cloonan et al., 2016). However, the effects of these specific digital learning and assessment tools are unknown.

It is unclear whether all computer-based formative assessments can influence student achievement and, as part of the steps of formative assessment by Sadler (1989), close the gap for students. The study by Clemens et al. (2015) investigated the effects of a computer-based adaptive formative test for 71 students in kindergarten and Grade 1. The researchers found that the test to be a statistically significant predictor of end-of-year reading-related skills. This article was the only one of its kind in the search from 2014 to 2019. Further research is needed to evaluate the ability of computer-based formative tests to influence student achievement.

The Istation Program

Researchers found that the Istation formative assessment program improves student achievement on summative assessments in varied locations. In a mixed methods study with 98 third-grade children, Luo et al. (2017) found a strong correlation between Istation individualized formative assessment and improved reading scores on the STAR; they recommended further research with larger samples, especially if controlled for extraneous factors. Putman (2016) found a statistically significant relationship between the Istation reading program and kindergarten literacy achievement and indicated a need to conduct additional studies with other grade levels and locations. Results from a study by Patarapichayatham et al. (2014) found that Istation served as a predictor of how students perform on the STAAR reading test; the authors suggested that additional research expand upon these findings. More recent studies found that Istation scores serve

as a predictor of scores on other state summative tests and support the need for additional replication and a study that further isolates Istation's effect on student preparedness for summative exams (Campbell et al., 2018; Patarapichayatham, 2018). Marin (2015) conducted a reading improvement study with Texas third graders (n = 102) and suggested that Istation's effectiveness as appropriate formative assessment should be investigated in other grade levels and locations. Thus far, no research evaluates the influence of the Istation program on the PARCC summative assessment. Furthermore, no studies are focused on the influence of the program on upper elementary students.

Conclusions

Formative assessment is widely accepted in the education community. Despite copious amounts of research in the formative assessment realm—including more recent online programs, those specific to reading, and both—the connection between specific formative assessments such as the Istation program on summative assessment remains unknown. One major theme in the literature is that formative assessment is a positive practice that is welcomed in most classrooms. A second theme is digital or computer-based tools are abundant in the education realm. Though computer-based formative assessments show some research-based effectiveness, specific programs' relationship to summative assessments are unclear. Istation is one of many programs claiming to help teachers and students understand where they are throughout the year as a way to improve student achievement. This study may contribute to the existing research about formative assessment and its relation to summative assessment and potentially help narrow the gap in practice using these tools.

Chapter 3: Research Method

The influence of specific formative assessment programs, like Istation, on student achievement remains unknown. Therefore, the purpose of this quantitative study was to investigate the influence of the Istation reading formative assessment program on student achievement as measured by the ELA PARCC summative assessment. In this chapter, I include the methods related to this quantitative study including the setting, research design and rationale, population, and data collection and procedures. The chapter concludes with a summary of the design, methodology, potential threats to validity, and ethical procedures.

Research Design and Rationale

To adequately respond to the first three research questions, I used a quantitative approach with a one-way within-subjects repeated measures ANOVA with one dependent variable (the ISIP score) and one independent variable measured at the three points in time the students took the Istation formative assessment: BOY, MOY, and EOY. To adequately respond to the second set of questions, I used another quantitative approach via a linear regression analysis with one dependent variable (the ELA PARCC) and MOY ISIP score as the independent variable. Since the literature did not indicate a need to investigate age or gender issues, I ran the ANOVA and regression for all students without these potentially confounding variables. In this section, I include information about the specific designs I chose and their relation to the research questions. To assess the influence of student performance on the Istation program over time and on students' achievement as evidenced by the summative exam, I used the research questions to find

whether there is a change in the ISIP scores for students over time and whether the Istation formative assessment scores relate to scores on the ELA PARCC.

Design

I used a quantitative design and the within-subjects one-way repeated measures ANOVA statistical test to examine whether the mean ISIP scores changed over time, thus indicating that students were learning with this formative assessment program. I chose a quantitative approach over qualitative or mixed methods because it is the most effective in identifying the extent to which the independent variables influence groups. The dependent variable is the Istation formative assessment program. The independent variable is the ISIP score from the three times the students took the Istation reading formative assessment program throughout the year (i.e., BOY, MOY, EOY).

I used a quantitative design and the linear regression model to better understand the relationship between formative and summative assessment. If the MOY ISIP score (independent variable) can predict the ELA PARCC score (dependent variable), there is evidence of the ability of the Istation reading formative assessment program to influence student achievement on the ELA PARCC summative assessment. This analysis will offer information about the predictive ability of the Istation program and offer benefits to teachers to better use the program throughout the year.

Within-Subjects ANOVA Rationale

According to Tabachnik and Fidell (2007), an ANOVA is used to determine whether there is a statistically significant difference among two or more means. A within-subjects ANOVA is appropriate when the evaluated means are from the same subjects

but “measured on different occasions” (Tabachnik & Fidell, 2007, p. 43). ANOVA provides an accurate method to determine the effects of an independent variable on one continuous dependent variable (Laerd Statistics, 2019). I investigated whether the Istation formative assessment program completes the steps outlined by Sadler (1989) by improving student skills over time. The same students completed all three measures. The ANOVA helped determine whether there is a change over time based on the three Istation assessment points in time (i.e., the independent variable). There were not issues with time or resource constraints as I pulled currently existing ex post facto data.

As formative assessment is widely used in the discipline of education, the repeated measures choice is consistent with previous literature investigating the effects of student performance on formative assessments related to summative assessment. Finding out whether the Istation program student performance influences achievement as measured by summative assessment adds to the knowledge of specific formative assessments on summative assessments and overall learning in the reading content area.

Linear Regression Rationale

The primary goal of the linear regression is to determine whether the independent variable can predict the dependent variable (Laerd Statistics, 2019). This analysis may reveal whether the linear regression line between two variables is statistically significant, how much variation in the dependent variable is explained by the independent variable, the direction and magnitude of the relationship, and the predicted values of the dependent variables based on values of the independent variable (Laerd Statistics, 2019). In this study, the linear regression will determine how much the dependent variable (i.e., ELA

PARCC) changes for a one-unit change in the independent variable (i.e., MOY ISIP). This analysis is the most appropriate choice because of the data involved and the need to assess the influence of the formative assessment on summative assessment.

Methodology

To find out whether the Istation reading program influences student achievement on the ELA PARCC, I quantitatively analyzed the influence of the Istation formative assessment on students who were or were not proficient on the PARCC summative assessment. I used archival data from the year all teachers were required to have students use the program with fidelity and take the three assessments (i.e., BOY, MOY, EOY). I used the student scores from the ELA PARCC exam taken in Spring 2018. In this section, I review the sampling procedures for data collection, the instrumentation, and the data analysis plan.

Population Selection and Sampling Procedures

The population for the archived data set for this study consisted of students from a Title I school in Grades 3-6 in the Southwestern United States ($N = 281$). In this study, I included all students who took all three Istation formative assessments (i.e., BOY, MOY, EOY) and the ELA PARCC assessment. There were four general education classrooms in each grade with about 20 students in each class. The number of students in each class fluctuated throughout the year due to late start and early withdrawals. Some students also attended special education services throughout the day. If a small group setting was listed in the accommodations section of the Individualized Education Plan, these students went to a different space for testing. District policy is that all students enrolled in school at

HES during the months of August and/or September take the Istation BOY, those enrolled in January take the MOY, and those enrolled in May take the EOY. The teachers and administrators gave the ELA PARCC test in April and/or May depending on student absences. Each student enrolled in school those months also took the ELA PARCC summative exam. If a student was absent, they took the PARCC test on a make-up testing day. Any students who took the alternative assessment based on their Individualized Education Plans were excluded from this group.

Sampling Strategy and Size. Quantitative research requires strategies for a sample to ensure proper analysis of the data. According to Faul, Erdfelder, Lang, and Buchner's (2007) G*Power software, the calculated power (1-β error probability) for the desired sample in this study was 0.903. The post hoc achieved power of 0.903 was numerically greater than the threshold value of 0.80, indicating that this analysis required a minimum sample size of 73 students (i.e., 0.903 = the achieved power for this statistical test at the parameters of effect size equaling 0.15 [medium], alpha at 0.05, total sample size equaling 73, and number of predictors equaling 1). Thus, for these statistical parameters, there would be sufficient power to support the analysis results if the data sample included at least 73 complete student data sets.

Sampling Frame. All students who were enrolled at HES for all three Istation measures and the ELA PARCC in August/September (BOY), January (MOY), May (EOY) for Istation and April/May for ELA PARCC assessment for the 2017–2018 school year had data sets that were potentially included in this study. As there were 281 students enrolled at HES for the indicated dates, this study potentially included data from 281

students. Alternately, students were excluded if they were not enrolled for all four tests. This process eliminated data from students who either enrolled after September 30 or who withdrew from the school before taking all the exams.

Archival Data

As I used archived ex post facto data, there were not participants in this study. I collected the deidentified archival data from the 2017–2018 Istation reading formative assessment program and the same students' scores on the culminating ELA PARCC assessment at HES in Grades 3-6 administered in 2018. All grades were included together as, according to the literature, there was no need to compare grade levels. The principal of HES agreed to print all data needed for this study.

I employed a quantitative ex post facto design to investigate the effects of the Istation reading formative assessment program on student achievement as measured by the ELA PARCC. Performance measures came from archival data reported for Istation and PARCC. Each student in Grades 3-6 used Istation reading at least 25 minutes a week. Student data and progress are automatically collated into a student Istation report each month called the Individual Development Summary on the Istation website. Following each monthly formative assessment, the Istation's computer-based program informs the student where they are regarding reading comprehension, understanding vocabulary, and spelling achievement, and individualizes tasks to close the gap and move the student toward achieving grade-level skills.

I collected the data from the BOY, MOY, and EOY because these were the data for this term and available to the principal and the state from 2017–2018. I included each

child who took all three Istation reading formative assessments (i.e., BOY, MOY, EOY) and the ELA PARCC summative exam for the 2017–2018 school year. Though most students enrolled throughout the school year had a complete set of data, some students were not included due to date of enrollment and/or withdrawal.

Access and Permissions. The beginning, middle, and end of year data are stored on the Istation's website with administrator access. At the end of the school term, student achievement is summatively assessed with the ELA PARCC assessment. The PARCC data are stored on the principal's computer and require prior approval to view. I obtained superintendent and principal approval to access and use these data for my study. The superintendent granted permission for the access and analysis of the data. The school's principal agreed to print and de-identify all the data needed for this study.

Instrumentation and Operationalization of Constructs

As the researcher of a quantitative study using archival data, I did not select an instrument to gather data. In this section I describe the design of the Istation reading formative assessment program, founded in 1998, as it is used by HES (see Owen, 2016). This program is appropriate to the study because it is used by all teachers at HES and in all schools in the state from kindergarten through Grade 2. Though it is optional, the superintendent uses budget funds to continue use of the program for Grades 3-6. The information gained from the study may inform administration about future use.

Istation. According to the Texas Success Training (2013), the Istation program was founded in 1998 and has four components: (a) assessments that are individualized and automatically available in about 30 minutes or less, (b) instruction, (c) reports that

provide regular data, and (d) teacher tools. Assessments are categorized as Early Reading (PK-3), consisting of phonemic awareness, vocabulary, and letter knowledge. The program has an additional comprehension component for kindergarten. In first and second grades, the program assesses and teaches alphabetic decoding. Third grade curriculum covers spelling, vocabulary, connected text fluency, and comprehension. The advanced assessment is for Grades 4 through 8 and includes "Word Analysis (Orthographic Representation), Fluency (Text Fluency Maze – 2 minutes 30 seconds – cloze passage), Vocabulary (General and Content), and Comprehension (Main Idea, Inference, Cause and Effect, and Critical Judgment)" (Texas Success Training, 2013, p. 1). Teacher tools contain over 2,000 lessons for paper practice and smart boards that are leveled according to difficulty (Texas Success Training, 2013).

Istation's designers created this assessment and instructional program to determine whether students are building the skills needed to become proficient readers and ultimately help them close the gap in knowledge from where they are to where they need to be. According to the Texas Success Training (2013), the recommended usage for each session is 25 minutes. As students advance in the program, monitoring takes place. It is recommended that the extra time for Tier 1 take place at least one time a week for 30 minutes. For students in Tier 2, an additional time should be provided, which is recommended to be at least 2 days a week for 30 minutes each time. Students in Tier 3 require at least 3 days during a week with at least 30 minutes or more each time for a total of 90 minutes or more. Instructional sessions include 12 cycles. The cycles contain Earth science, some mathematics, and science content passages, with teachers able to

choose between narrative or expository text. In 2013, the designers added new materials that included over 15 cycles after the ISIP assessment, which include the 6 + 1 traits, persuasive essays, expository, and narrative materials in the writing rules. Social studies skills, tailored for older students who are more advanced, are included in a lesson called Timeless Tales.

Previous Publications Involving Istation. Though recent research included the use of the Istation program and its influence or ability to predict summative assessment scores, these studies did not include reliability and validity values of the program. There are currently no published validity measures. I contacted the designers of the program via email to further investigate this issue.

Operationalization of Variables. To properly measure the variables for the research questions in this study related to the one-way repeated measures ANOVA, I accessed data on the dependent (ISIP score) and three independent variables (i.e., three points in time the students took the Istation formative assessment). To properly measure the variables for the questions related to the linear regression, I accessed data on the dependent (ELA PARCC score) and the independent variable (MOY ISIP score) to test predictability. These data came from existing district data sets. In this section, I provide detail on the measure for each variable, how the variable score is calculated, and what the score represents.

ELA PARCC Assessment. The dependent variable in this study was student scores on the ELA PARCC assessment. The PARCC is a yearly summative assessment that students take in Grades 3-8 in both ELA and mathematics ((New Jersey Department

of Education, 2015). In this study, the ELA assessment data were used to determine the predictive ability or influence of the Istation formative assessment program. The PARCC ELA Reading Assessment measures students' abilities in reading as they relate to the grade-level standards and text complexity band for the grade level. The PARCC Communications Team (2014) stated:

> The PARCC assessments are designed to measure the academic standards in the English language arts/literacy and mathematics Common Core State Standards. The standards are the constructs that identify what students should be taught and learn at each grade level so that by the time they graduate from high school they have the reading, writing, and mathematical knowledge and skills needed to succeed in college and/or jobs with career potential. (para. 5)

The assessment consists of reading multiple complex grade-level passages in both fiction and nonfiction. Passages may also contain video and audio as decided upon within certain standards. Student comprehension is assessed through multiple question types including multiple choice, constructed response, and drag and drop. Students also take information from the reading passages to construct extended-response writing in which textual evidence is used to demonstrate an understanding of the text.

The PARCC Measure. PARCC reports student performance in five levels based on the knowledge, skills, and practices aligned to the grade-level CCSS that students demonstrate within the assessment. The performance levels include:

- Level 1: Did not yet meet expectations (650–699)
- Level 2: Partially met expectations (700–724)

- Level 3: Approached expectations (725–749)
- Level 4: Met expectations (750–809)
- Level 5: Exceeded expectations (810–850)

Students who attain a Level 4 or 5 are considered proficient (i.e., values range from 750–850). Proficient students meet grade-level expectations or what the PARCC Communications Team (2014) described as what typical students at each level should be able to demonstrate based on their command of grade-level standards.

The PARCC assessment provides educators with data relating to a student's abilities as compared to grade-level expectations in the reading areas of literary text, informational text, writing, and vocabulary (PARCC Communications Team, 2014). Pearson (2017) stated that gathering construct validity evidence for PARCC is embedded in the process by which the PARCC assessment content is developed and validated. At each step, "the states involved hundreds of educators, assessment experts, and bias and sensitivity experts in review of text, items and tasks for accuracy, appropriateness, alignment to the instructional standards" (Pearson, 2017, p. 115). The average internal consistency reliability for the ELA assessment is a range of 0.91–0.93 (Pearson, 2017). The PARCC assessment received an excellent match to the Common Core State Standards criteria for content in ELA (Pearson, 2017). The PARCC assessments meet or exceed the depth and complexity required by the criteria through a variety of item types that are generally of high quality, measuring grade-level standards (Pearson, 2017). The administrators at HES were state mandated to have each student participate in the PARCC assessment for the 2017–2018 school year.

***PARCC Administration*.** The teachers and administrators administered the test over a 3-day period, and students received up to 90 minutes per session. These administration days were not always consecutive as the school environment allows for only one school level or grade to take the test at a time. There were up to 12 questions per session, with three texts to read and one written component based on the reading. The ELA PARCC assessment evaluated grade-level skills that align with the CCSS. Students took this assessment during the assessment period, which ran from mid-April to the beginning of May 2018.

***Istation ISIP scores*.** The independent variables in this study were measured by student scores from the Istation program's reading formative assessments that were completed at three intervals: the beginning, middle, and end of the 2017–2018 school year. The ISIP score measures student overall reading skills, and the student's score was the measured independent variable (i.e., values for Grade 3 range from 196–293; Grade 4, 1,330–2,200; Grade 5, 1,600–2,600). New assessments are automatically administered on the first day of each month and the level of difficulty increases systematically through the Istation program. All Istation evaluations are timed, and the program provides a report of students who had inactive periods or terminated the program. For instance, one assessment affords students 4 minutes and 30 seconds to read a passage. If a student does not finish reading the passage, then the student is moved on. Students cannot return to the passage to seek the answers; rather, they can only refer to the instruction sessions. The school administration reports data on the pre-, mid-, and posttest. The three independent variables represent how the student performed on the assessment at three points in time:

the beginning of the year (BOY) in August or September 2017, middle of the year (MOY) in January 2018, and end of year (EOY) in May 2018.

Data Analysis Plan

After gathering the data, I analyzed them following the steps described in this detailed analysis plan that includes the identification of the software and processes. Next, I list the specific research questions and the hypotheses, an explanation of data cleaning and screening procedures, the statistical tests used to test the hypotheses, and how the results were interpreted.

Software. I imported the data into the IBM Statistical Package for the Software Sciences (SPSS) 25 software. I manually entered the data for each variable in its own column. After entering the data, I used the SPSS program to run general linear model repeated measure and regression procedures.

Data Cleaning and Screening Procedures. After receiving the deidentified data, I cross-referenced the data by each subject and ensured the data were placed correctly in each box. This process was necessary and is called *cleaning and screening*. Completing this task helped to eliminate the students who did not take all the assessments and were, therefore, not included in the sample. These data were screened by cross-referencing which students' data sets were incomplete because they were missing one or more of the data points from either of the three Istation assessments and/or the ELA PARCC Assessment. This process was accomplished by first working from the list of students who took the ELA PARCC and assigning a number starting at 1 to the beginning of that row. Then, I gathered the rest of that student's scores from the BOY, MOY, and EOY and

entered those data into the same row. Finally, if a student did not have all data points, I deleted that entire row and made a notation on the original form. After the cleaning and screening process was complete, to ensure accuracy, I went over the entire list a second time. I found no errors.

Research Questions and Hypotheses. The purpose of this study was to investigate the influence of the Istation reading formative assessment program on student achievement as measured by the ELA PARCC assessment. To complete this investigation, I used quantitative analysis with research questions related to the change in student scores on the Istation program over time and the relationship between the Istation formative assessment program and the ELA PARCC. I included null and alternate hypotheses. The research questions are as follows:

RQ1: Are there statistically significant differences in ISIP scores for Grade 3 before (BOY), during (MOY), and after (EOY) participation in the Istation reading program?

$H_0$1: There are no statistically significant differences in ISIP scores for Grade 3 before (BOY), during (MOY), and after (EOY) participation in the Istation reading program.

H_a1: There are statistically significant differences in ISIP scores for Grade 3 before (BOY), during (MOY), and after (EOY) participation in the Istation reading program.

RQ2: Are there statistically significant differences in ISIP scores for Grade 4 before (BOY), during (MOY), and after (EOY) participation in the Istation reading program?

$H_0$2: There are no statistically significant differences in ISIP scores for Grade 4 before (BOY), during (MOY), and after (EOY) participation in the Istation reading program.

H_a2: There are statistically significant differences in ISIP scores for Grade 4 before (BOY), during (MOY), and after (EOY) participation in the Istation reading program.

RQ3: Are there statistically significant differences in ISIP scores for Grade 5 before (BOY), during (MOY), and after (EOY) participation in the Istation reading program?

$H_0$3: There are no statistically significant differences in ISIP scores for Grade 5 before (BOY), during (MOY), and after (EOY) participation in the Istation reading program.

H_a3: There are statistically significant differences in ISIP scores for Grade 5 before (BOY), during (MOY), and after (EOY) participation in the Istation reading program.

RQ4: To what extent do Istation formative assessment scores (MOY) predict literacy performance as measured by the ELA PARCC summative assessment for Grade 3?

$H_0$4: The Istation formative assessment does not predict literacy performance as measured by the ELA PARCC summative assessment for Grade 3.

H_a4: The Istation formative assessment does predict literacy performance as measured by the ELA PARCC summative assessment for Grade 3.

RQ5: To what extent do Istation formative assessment scores (MOY) predict literacy performance as measured by the ELA PARCC summative assessment for Grade 4?

$H_0$5: The Istation formative assessment does not predict literacy performance as measured by the ELA PARCC summative assessment for Grade 4.

H_a5: The Istation formative assessment does predict literacy performance as measured by the ELA PARCC summative assessment for Grade 4.

RQ6: To what extent do Istation formative assessment scores (MOY) predict literacy performance as measured by the ELA PARCC summative assessment for Grade 5?

$H_0$6: The Istation formative assessment does not predict literacy performance as measured by the ELA PARCC summative assessment for Grade 5.

H_a6: The Istation formative assessment does predict literacy performance as measured by the ELA PARCC summative assessment for Grade 5.

ANOVA assumptions. For a one-way repeated measures ANOVA to be an appropriate analysis, the data set must meet five assumptions. The first two require that (1) the dependent variable must contain continuous data and (2) the independent variable contains at least two categorical levels (Laerd Statistics, 2019). Based on the data set for

this study, the dependent variable was the ISIP score, which is tabulated on a continuous scale. The within-subjects independent variable has three levels, representing the points at which the students took the Istation formative assessment. Therefore, the data set should meet these first two assumptions.

The next three assumptions involve the nature of the data. These assumptions require that (3) there are no significant outliers existing in any cell of the data set, (4) the dependent variable is normally distributed, and (5) the variance of the dependent variable approximates the same in each subject, also known as sphericity (Laerd Statistics, 2019). When each of these assumptions is met, it is appropriate to run an ANOVA. Tests are necessary to assess whether these assumptions are met with the existing data set.

Outliers. To address the third assumption regarding potential significant outliers, I first needed to identify any outliers by interpreting boxplots. If the outliers existed due to a data entry error, I simply fixed this issue by correcting the data and reran all the assumption tests. Then, Laerd Statistics (2019) suggested checking for measurement error, which relates to an equipment malfunction or out-of-range values. These issues can also be addressed by correcting the data. Once these two errors are checked and addressed, a third step may be to decide to keep or remove the outliers.

If necessary, mitigating outliers can be handled in one of two ways. I can keep the outliers by transforming the dependent variable to include the outliers or running the one-way repeated measures with and without the outliers to see whether the results are substantially affected (Laerd Statistics, 2019). The second option is to remove the outliers. However, proper justification will be necessary because this option is typically

viewed as a last resort (Laerd Statistics, 2019). Once these assumptions are met, addressing the normal distribution of the dependent variable is next.

Normal distribution. Normality of the dependent variable is necessary to establish statistical significance (Laerd Statistics, 2019). The most common method to do so is to run a Shapiro-Wilk test for normality. If data are normally distributed, the significance value should be more than 0.05, and I will then move on to assumption five (Laerd Statistics, 2019). However, if the data violate the assumption of normality due to a significance value of less than 0.05, further identifying of outliers is necessary as discussed for Assumption 3 (Laerd Statistics, 2019). Once Assumption 4 is met, it is appropriate to move forward and address Assumption 5.

Sphericity. Finally, Assumption 5 relates to equal variance of the dependent variable, known as sphericity. According to Tabachnick and Fidell (2017), *sphericity* is defined as measuring the correlation scores between the dependent variables that occur over time to determine whether they are similar. A violation of sphericity can lead to invalid results (Laerd Statistics, 2019). Since correlations are more likely to be similar between variables that are measured closer together in time, Mauchly's test can determine sphericity. Using Mauchly's test statistic, sphericity can be either significant ($p < .05$) or nonsignificant ($p > .05$). If there is a violation of sphericity, additional statistical tests can be used to correct for the error (Tabachnick & Fidell, 2007). Once the data meet all the five assumptions, it is appropriate to run the ANOVA.

***Interpreting results*.** The primary goal of the one-way repeated measures ANOVA is to determine whether there are any statistically significant differences

between the means of three or more levels of a within-subjects factor (Laerd Statistics, 2019). If there is a statistically significant one-way interaction on the dependent variable, I will then interpret results of the Sig values using the output table titled Tests of Within-Subjects Effects. If the results in the Sig column satisfy $p \leq .05$, there is a statistically significant interaction. Conversely, if the Sig value shows $p \geq .05$, there is not a statistically significant interaction. I will then run post hoc tests to further examine the difference in means from and to specific time points. These analyses will show whether there were statistically significant differences in means between all the time points (i.e., BOY, MOY, EOY).

Linear Regression Assumptions. For a simple linear regression to be an appropriate analysis, the data set must meet seven assumptions. The first two are that the data must have a (1) continuous dependent variable and (2) continuous independent variable (Laerd Statistics, 2019). Based on the data set, the dependent variable is the ELA PARCC raw score, which is tabulated on a continuous scale. The MOY ISIP score represents the independent variable, also tabulated on a continuous scale. Therefore, the data should meet the first two assumptions.

The next three assumptions involve the nature of the data and can be evaluated by following the linear regression procedure in the SPSS program. These assumptions include the following: (3) there is a linear relationship between the two variables; (4) the data have independence of observations; (5) there are no significant outliers in the data set; (6) the data have homoscedasticity; and (7) the residuals (errors) of the regression line are approximately normally distributed (Laerd Statistics, 2019). Tests are necessary

to assess whether these assumptions are met with the existing data set. When each of these assumptions is met, it is appropriate to run a linear regression analysis.

Linear relationship. To address the third assumption regarding a linear relationship between the two variables, I used the SPSS program to run a scatter plot of the dependent variable plotted against the independent variable to see if a linear relationship exists (Laerd Statistics, 2019). If these results render a straight line, a linear relationship exists. However, if these data violate this assumption, they can either be transformed to coax them into a linear regression or another analysis can be done, such as a polynomial or nonlinear regression. Once the data meet the linear regression assumption, it is appropriate to move on to assessing independence of observations.

Independence of observation. The fourth assumption involves independence of observations. According to Laerd Statistics (2019), one residual cannot provide information about another residual. Thus, independence means the observations cannot overlap. The way to assess independence of observations is by reviewing the data from the Durbin-Watson test (Laerd Statistics, 2019). The Durbin-Watson results range from 0 to 4. If the results of the analysis are close to 2, this assumption is considered met and the researcher can then move on to Assumption 5.

Outliers. To address the fifth assumption, one must check for outliers. These outliers can be leverage points or influential cases. According to Laerd Statistics (2019), these outliers can all be considered unusual points that significantly differ from the usual trend of the data points. Cook's distances analysis can be used to identify these outliers. Any result greater than one requires further inspection. After running this analysis, a

separate Cook's variable will appear in the data set. Further review of these outliers can be conducted to see whether removing the outliers improves the regression line and reduces the influence of a specific data point.

Homoscedasticity. The sixth assumption involves homoscedasticity. This assumption can be evaluated by inspecting the plot of residual values against predicted values (Laerd Statistics, 2019). If there is homoscedasticity, the residuals will be equal across the predicted values (Laerd Statistics, 2019). This means there is a constant spread and no pattern in the data points, the homoscedasticity assumption is met, and it is appropriate to move on to assess the next assumption.

However, if these data violate the homoscedasticity assumption, there are two ways to resolve this issue. Transforming the data could remove heteroscedasticity (Laerd Statistics, 2019). A second way to eliminate this concern with the data is to run a weighted least-squares regression, a regression with robust standard errors, or a robust regression (Laerd Statistics, 2019). Once these data meet the homoscedasticity assumption, it is necessary to move on to the next assumption.

Normal distribution. Finally, to address the seventh assumption, it is necessary to evaluate whether the residuals (errors) of the regression line are approximately normally distributed. Two methods can be used to make this determination: inspecting a histogram or a normal probability plot (Laerd Statistics, 2019). If these data do not appear normally distributed, it would be necessary to transform the dependent and possibly the independent variable (Laerd Statistics, 2019). As this is the final assumption, once these data are approximately normally distributed, it is appropriate to run the linear regression.

***Interpreting results*.** The primary goal of the linear regression is to determine whether the independent variable can predict the dependent variable. It is effective with continuous data and strives to create regression coefficients (Tabachnik & Fidell, 2007). In this study, a simple linear regression was performed to create an equation that would best quantify the relationship between the independent variable, ISIP MOY reading formative assessment score, and the ELA PARCC summative assessment score.

In this study, the linear regression determined how much the dependent variable (ELA PARCC) changes for a 1-unit change in the independent variable (MOY ISIP). I performed a quantitative analysis using SPSS Version 25.0, a standard confidence interval of 95%, and an alpha of 0.05. The conventional medium effect size of 0.25 was used (Cohen, 1992). I interpreted the results using the output tables titled Model Summary and ANOVA. The analysis requires evaluation of the effect size via adjusted R^2, whether the coefficients show a linear relationship between the two variables, and the ability of the independent variable (Istation program, ISIP score) to predict the dependent variable (ELA PARCC score).

Threats to Validity

The main goal of this study was to determine whether there is a relationship between the Istation reading formative assessment and the PARCC summative assessment. According to the Standards for Educational and Psychological Testing, validity is "the degree to which evidence and theory support the interpretation of test scores for proposed uses of tests" (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education,

1999, p. 11). In this next section, I explain the potential threats to external, internal, and construct validity as well as how to address them.

External Validity

External validity involves generalizability. According to Campbell and Stanley (1963), factors that jeopardize external validity include interaction effect of testing, interaction effects of selection bias and the experimental variable, reactive effects of experimental arrangements, and multiple treatment inference. Because I used ex post facto data, there was no treatment or experimental variable. The quantitative design I chose eliminated these threats to external validity.

External validity can be defined as the degree to the generalization of results to other populations and locations (Shadish, Cook, & Campbell, 2002). Due to the specific population and location chosen, one potential threat could be the sample used since it involves the data from the population of one Title I school in one district in a rural area from one southwest state. Therefore, the results may only be generalizable to students in areas with similar socioeconomic status and type of location.

Internal Validity

Internal validity relates to the ability of the design to actually test the hypotheses I intended it to test. Campbell and Stanley (1963) listed eight classes of extraneous variables that could produce confounding effects on the experimental design. These variables include history, maturation, testing, instrumentation, statistical regression, selection bias, experimental mortality, and selection maturation (Campbell & Stanley,

1963). Most of these threats are eliminated with the use of ex post facto data since there is not an experimental design.

However, one potential threat to internal validity could be under the testing category because external factors could have influenced student performance on the exams listed as independent and dependent variables. The teachers typically proctor their own students' tests with administrator supervision in two locations based on the testing schedule: the technology lab and library. Both locations have student restrooms nearby with some distractions that are not documented.

History is a potential second threat to internal validity for this study. History refers to events that could have happened in between the assessments (Campbell & Stanley, 1963). Students may have experienced issues in their home or family that could have affected their performance and achievement on the Istation formative tests or proficiency on the ELA PARCC test. Since these data are not recorded, the outside effects on the measures are not included in the study.

Construct Validity

Construct validity refers to the concepts of the study. Construct validity reflects whether the implemented intervention is the intervention that was intended to be implemented and whether the outcomes were measured as it was intended to be measured (Trochim, Donnelly, & Arora, 2016). One potential threat to construct validity is an inadequate explanation of the constructs. To reduce this threat, I have clearly and operationally defined each construct involved in the study.

Ethical Procedures

Ethical issues must be addressed in quantitative research to ensure that potential harm is minimized (Creswell, 2009). Since I obtained only archival ex post facto data for this study, potential harm was eliminated as it did not involve any interactions with or observations of human participants. Prior to beginning this study, I obtained Institutional Review Board (IRB) approval. Because the data that were analyzed are existing from previous assessments, there was no threat of ethical issues related to recruitment or incentives. Therefore, the ethical procedures listed in this section relate to the treatment of archival data and two potential ethical issues related to the data gathered.

Treatment of Archival Data

The ex post facto data I received was de-identified. During analysis, I was the only one with access to the data. The principal could not grant me access to an online version or Excel spreadsheet, so she printed all the data by grade level. I stored these data in a locked filing cabinet. Upon completion of the analyses, I shredded all documents in the school office.

Other Ethical Issues

One potential ethical issue was working in the same location the study took place. However, since I used ex post facto data, this issue was not a conflict of interest. The deidentified data prevented me from knowing how specific students scored on the assessments. Another potential issue was that I am one of the teachers that conducted this assessment in the 2017–2018 year. However, upon initial gathering, there was no plan of using this data, and I only received information on the individual student, so I was not

privy to which teacher each student had that year. This alleviated potential bias as I did not know which students on the spreadsheet were a part of my classroom at that time.

Summary

In this quantitative study, I analyzed archival data to investigate the influence of student performance on the Istation reading formative assessment on achievement as measured by the ELA PARCC assessment for Grades 3-6 at HES. This chapter included justification for using archived data for both a one-way repeated measures ANOVA and a linear regression design. Next, I presented the sample and data collection procedures. In the data analysis section, I included the procedures for running statistical tests and the ways to check for error. Further, I examined potential threats to the study's internal, external, and construct validity and explained how I plan to mitigate the threats. Finally, I discussed ethical procedures and how to address each of them. I received approval of the proposal from the IRB and received the following approval number: 04-01-20-06668265. In Chapters 4 and 5, I provide detailed description of how I executed the data collection and statistical analysis of data.

Chapter 4: Results

The purpose of this quantitative study was to investigate the influence of the Istation reading formative assessment program on student achievement as measured by the ELA PARCC summative assessment. Therefore, the research questions I investigated sought to find out whether there was a change in the reading ISIP score over time and whether there was a relationship between Istation scores and performance on the ELA PARCC. The research questions in this study were as follows:

RQ1: Are there statistically significant differences in ISIP scores for Grade 3 before (BOY), during (MOY), and after (EOY) participation in the Istation reading program?

$H_0$1: There are no statistically significant differences in ISIP scores for Grade 3 before (BOY), during (MOY), and after (EOY) participation in the Istation reading program.

H_a1: There are statistically significant differences in ISIP scores for Grade 3 before (BOY), during (MOY), and after (EOY) participation in the Istation reading program.

RQ2: Are there statistically significant differences in ISIP scores for Grade 4 before (BOY), during (MOY), and after (EOY) participation in the Istation reading program?

$H_0$2: There are no statistically significant differences in ISIP scores for Grade 4 before (BOY), during (MOY), and after (EOY) participation in the Istation reading program.

H_a2: There are statistically significant differences in ISIP scores for Grade 4 before (BOY), during (MOY), and after (EOY) participation in the Istation reading program.

RQ3: Are there statistically significant differences in ISIP scores for Grade 5 before (BOY), during (MOY), and after (EOY) participation in the Istation reading program?

H_03: There are no statistically significant differences in ISIP scores for Grade 5 before (BOY), during (MOY), and after (EOY) participation in the Istation reading program.

H_a3: There are statistically significant differences in ISIP scores for Grade 5 before (BOY), during (MOY), and after (EOY) participation in the Istation reading program.

RQ4: To what extent do Istation formative assessment scores (MOY) predict literacy performance as measured by the ELA PARCC summative assessment for Grade 3?

H_04: The Istation formative assessment does not predict literacy performance as measured by the ELA PARCC summative assessment for Grade 3.

H_a4: The Istation formative assessment does predict literacy performance as measured by the ELA PARCC summative assessment for Grade 3.

RQ5: To what extent do Istation formative assessment scores (MOY) predict literacy performance as measured by the ELA PARCC summative assessment for Grade 4?

$H_0$5: The Istation formative assessment does not predict literacy performance as measured by the ELA PARCC summative assessment for Grade 4.

H_a5: The Istation formative assessment does predict literacy performance as measured by the ELA PARCC summative assessment for Grade 4.

RQ6: To what extent do Istation formative assessment scores (MOY) predict literacy performance as measured by the ELA PARCC summative assessment for Grade 5?

$H_0$6: The Istation formative assessment does not predict literacy performance as measured by the ELA PARCC summative assessment for Grade 5.

H_a6: The Istation formative assessment does predict literacy performance as measured by the ELA PARCC summative assessment for Grade 5.

Each of these research questions included grade-level specific and complete sets of data based on student scores from the 2017–2018 school year. RQ 1-3 involve the Istation program scores and whether a one-way repeated measures ANOVA indicated that the scores change over time. Questions 4-6 involve the relationship between the Istation formative assessment program and its ability to predict and influence the ELA PARCC summative assessment. The rest of this chapter includes data collection, results from both statistical tests for Grades 3-5 and an analysis of these results.

Data Collection

After receiving IRB approval (#04-01-20-0668265), I collected all data for this study following the methodology and best practice. As I used archived ex post facto data, I did not have participants in this study. I collected the following de-identified data: (a)

Grades 3-6 student formative assessment data from the Istation reading program and (b) summative ELA PARCC data from the same students. This section includes a narration of the data retrieval process and discrepancies from my data collection plan. Additionally, it includes a baseline description of the data, valuable for determining the validity and representativeness of the sample.

Retrieval Process

Upon validating IRB approval to gather data for this study, the HES principal provided me a printed copy of the data set for this study ex post facto in a password-protected spreadsheet. The Istation data set included 2017–2018 student scores from the beginning (BOY), middle (MOY), and end of the school year (EOY) for 270 students.

The data included in this study were the student's Istation formative assessment scores from the 2017–2018 school year. These were the existing data for this term and available to the principal and the state. I included student data sets for each student who completed a pre-, midyear, and postformative reading assessment in Istation and the ELA PARCC summative exam for the 2017–2018 school year. Though many students enrolled throughout the school year had a complete set of data, some were not included due to date of enrollment and/or withdrawal. Due to the data available, there was one significant discrepancy from the initial proposal.

Discrepancies in Data From the Plan

Although data were gathered according to the data collection plan, upon entering the data from a spreadsheet provided by the principal, I noted that none of the sixth-grade students had a BOY score. As the inaugural Istation assessment was in October of 2017,

the sixth-grade students from the data set did not have a BOY score—the pretest given at the beginning of the term. Therefore, all sixth-grade student data sets (n = 69) were incomplete and excluded, resulting in a data set of 201 rather than 270, representing students from Grades 3-5 rather than 3-6. Additionally, when I examined all Grades 3-5 data sets, another 26 were excluded. Therefore, from the 281 data sets initially retrieved, when I eliminated students with a missing data set or who unenrolled, the final sample was 175 Grade 3-5 students.

According to Faul et al. (2007) G*Power software, the post hoc achieved power (1-β error probability) for the sample in this study was 0.9991. The post hoc achieved power of 0.9991 is numerically greater than the threshold value of 0.80, indicating that this analysis has sufficient power to support the results with this statistical test at the parameters of effect size equaling 0.15 [medium], alpha at 0.05, total sample size equaling 175, and number of predictors equaling 1. Thus, for these statistical parameters, there was sufficient power to support the analysis results from the data sample of 175 complete student data sets.

One final deviation from the proposed data plan was that, because all students did not complete the ELA PARCC prior to the Istation EOY assessment, the MOY rather than the EOY was used as the predicting variable in determining the influence of ISIP scores on the ELA PARCC. As the MOY and EOY are the same type data, there were no negative effects on data analysis or meeting analysis assumptions created by this change. The discussion of data results reflects these deviations from the data plan.

Sample Description

The sample included 175 Grade 3-5 student data sets from a rural school in the Southwest region of the United States. As this school district was a Title I school, most students' families had a low socioeconomic status, and all students received free breakfast and lunch provided by the school. These data included scores from both male and female students, although gender was not coded in the data set since it was not a variable in this study. Table 1 shows the number of student data sets by grade level and the percentage of the sample.

Table 1

Total Number and Percentage of Student Data Sets by Grade Level

Grade level	Total	%
3	53	31
4	55	31
5	57	38
Total	175	100

HES is a Title I school in the southwest region of the United States. These students represented three different grade levels—Grades 3, 4, and 5. This sample (n = 175) represented 62% of the school's total population (N = 281). Since Grade 6 was eliminated (n = 69), the sample represented 83% of the population in Grades 3-5 (N = 212). Due to the number of complete data sets and the power (1-β error probability) for the desired sample in this study calculated at 0.903, the results can be generalized to the

overall population of students at HES in Grades 3-5. Furthermore, due to the diverse population of students at HES, the results can be generalized to other schools with similar populations.

Results

The following data, acquired from the archives at HES, informed the research questions for this study. Of the six research questions in this study, I analyzed three with a one-way repeated measures ANOVA and three with a linear regression. I conducted a one-way repeated measures ANOVA to examine whether there was a change in Grades 3-5 students' ISIP scores over time throughout the 2017–2018 school year—before, during, and after participation in the Istation reading program. The null hypotheses for RQ 1-3 stated that there are no statistically significant differences in ISIP scores before, during, and after participation in the Istation reading program. A one-way repeated measures ANOVA was appropriate because the teachers administered the Istation test three times throughout the school year.

The second set of questions, RQs 4-6, required a different type of quantitative analysis to examine the relationship between the ISIP scores and student performance on the ELA PARCC for each of Grades 3, 4, and 5. The null hypotheses stated that the MOY ISIP does not predict the ELA PARCC score. Linear regressions are appropriate to examine the extent of a relationship, if any, between two variables—an independent variable ~~a~~nd a dependent variable. Completing this analysis determined the influence of the Istation reading program on the ELA PARCC achievement.

Description of the Data

The ex post facto data from 175 HES students included all the Istation scores and the ELA PARCC score for each student that was enrolled at HES during the school year of 2017–2018. The data set originally included data from 281 students (de-identified and numbered) and included the grade level and, if taken, the BOY, MOY, EOY, and ELA PARCC scores. If the student was not present for one or more of the assessments, that space was left blank on the spreadsheet, and the student data set was excluded from the study. The ISIP score for each student was a numeral between 200 and 2,500. The ELA PARCC scores from 1 to 5 with the following criteria: did not yet meet expectations (1), partially met expectations (2), approached expectations (3), met expectations (4), and exceeded expectations (5). The state administration deemed students who scored a 1 through 3 as not proficient and a 4 or 5 as proficient.

This complete data set included scores for 175 students. Each of these students spent a minimum of 25 minutes on the reading component of the Istation program each week. Though the number of students fluctuated throughout the year, at the end of the year, the school had 281 students. However, the usable sample obtained for the purpose of this study included 175 students. Any student data sets from Grades 3-5, as well as Grade 6, were excluded if they were missing a score from one or more of the four tests required to have a complete data set. Data in Table 2 detail the number and percentage of Grade 3-6 students completing the ISIP at the beginning, middle, and end of the term.

Table 2

Number and Percentage of Students by Grade Level Completing the Istation ISIP at the Beginning (BOY), Middle (MOY), and End (EOY) of Year

	Total		BOY		MOY		EOY	
Grade	*N*	%	*N*	%	*N*	%	*N*	%
3	53	31	53	31	53	22	53	22
4	55	31	55	31	55	23	55	23
5	67	38	67	38	67	28	67	28
6	-		0		61	25	61	25
Total	175		175		236		236	

Note. BOY = beginning of year; MOY = middle of year; EOY = end of year

Statistical Assumptions for One-Way Repeated Measures ANOVA

For the one-way repeated measures ANOVA to be an appropriate analysis, five assumptions about the data must be met. According to Laerd Statistics (2019), verifying that a data set meets these assumptions is key to interpreting the validity of the results. However, it is common in real-world data for a data set to fail an assumption (Laerd Statistics, 2019). In these instances, it is essential to apply appropriate solutions and possibly further testing to overcome the violation of the assumption. The five assumptions for a one-way repeated measures ANOVA are as follows:

1. The dependent variable must contain continuous data.
2. The within-subjects independent variable must be categorical with three or more levels.

3. There must be no significant outliers existing in any level of the within-subjects factor.
4. The dependent variable must be normally distributed.
5. Known as sphericity, the variances of the differences between all combinations of levels of the within-subjects factor must be equal. (Laerd Statistics, 2019, "Assumptions I," para. 3)

Assumptions 1–2. According to the first assumption, the dependent variable in this study must contain continuous data (Laerd Statistics, 2019). In this study, the Istation formative assessment ISIP scores represent the dependent variable, which are tabulated on a continuous scale (i.e., values for Grade 3 range from 196–293; Grade 4, 1,330–2,200; Grade 5, 1,600–2,600). Since the dependent variable is tabulated on a continuous scale, this data set meets Assumption 1.

The second assumption states that the within-subjects factor (i.e., independent variable, the Istation reading formative assessment) contains at least three categorical levels (Laerd Statistics, 2019). For this study, the three levels of the independent variable represent the scores from the three times students took the Istation reading formative assessment throughout the year (i.e., BOY, MOY, EOY). As the scores were tabulated on a continuous scale and three levels exist, these data met Assumption 2. As the data met the first two assumptions, I continued to analyze data relative to Assumptions 3–5 that involve the nature of the data.

Assumption 3. The third assumption review consisted of checking for outliers. According to Laerd Statistics (2019), to determine the presence of outliers and normal

distribution in the data set, it is necessary to review the boxplots for the data set. To analyze the 175 data sets of ISIP scores, I used the Explore: Plots tab in SPSS to create boxplots for these data.

When analyzing outliers with boxplots, according to Laerd Statistics (2019), "Any data point that is more than 1.5 box lengths from the edge of their box is classified by SPSS Statistics as an outlier" ("Determining If You Have Outliers," para. 2). Those data points that are more than three box-lengths away from the edge of their box are considered extreme points and need to be further investigated (Laerd Statistics, 2019). The process for determining and problem-solving outliers is organized by research question.

RQ1 Outliers: Third grade ISIP scores. When analyzing the 53 data sets of third grade students' ISIP scores, I used the Explore: Plots tab to create boxplots for the data set. In the data set for Grade 3 (n = 53), there were two outliers: data points for students 13 and 41. Figure 1 includes the boxplots and evaluated outliers. When there are outliers in the data set as there were in this study, best practice requires an examination of the data set for (a) data entry errors, (b) measurement errors, or (c) genuinely unusual values (Laerd Statistics, 2019). Upon examination of these data, there were no entry or measurement errors. Therefore, the outliers were determined to be genuinely unusual values. With unusual values, best practice involves a process of determining whether the outlier should be kept or removed from the data set.

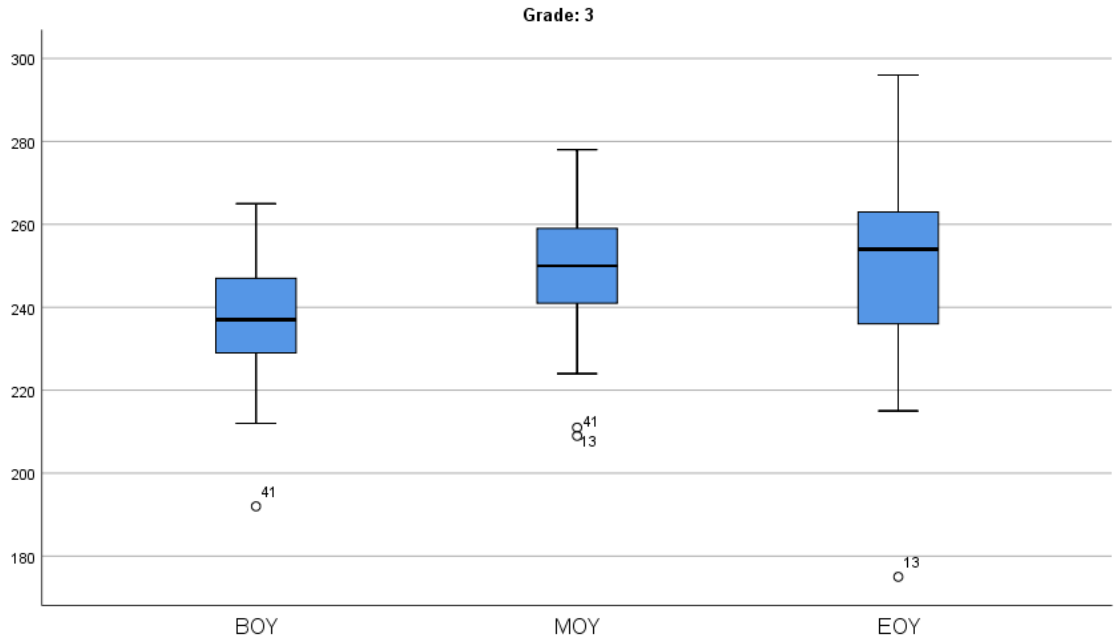

Figure 1. Boxplots for Grade 3.

There are four ways to resolve the problem of outliers. One can consider (a) using the nonparametric Friedman test, (b) modifying the outliers, (c) transforming the dependent variable, or (d) including the outlier in the analysis anyway (Laerd Statistics, 2019). As seen in Figure 1, none of the outliers were extreme points because, according to Laerd Statistics (2019), extreme points are more than three box-lengths away from the edge of their box. Since these outliers were not extreme, I chose to keep the outliers in the analyses. I calculated a one-way ANOVA with and without the outliers, compared the results, and found that the outliers had no effect on the analysis. The results were essentially the same (i.e., no statistically significant difference). Once these identified outliers were addressed, I moved on to Assumption 4 to determine the normality of the data.

RQ2 Outliers: Fourth grade ISIP scores. When analyzing the 55 data sets of fourth grade students' ISIP scores, I used the Explore: Plots tab to create boxplots for the data set. In the data set for Grade 4 (n = 55), these data rendered outliers for students 61, 78, 80, 82, and 85. Figure 2 includes the boxplots and evaluated outliers. Upon examination of these data, there were no entry or measurement errors. Therefore, the outliers were determined to be genuinely unusual values. With unusual values, best practice involves a process of determining whether the outlier should be kept or removed from the data set.

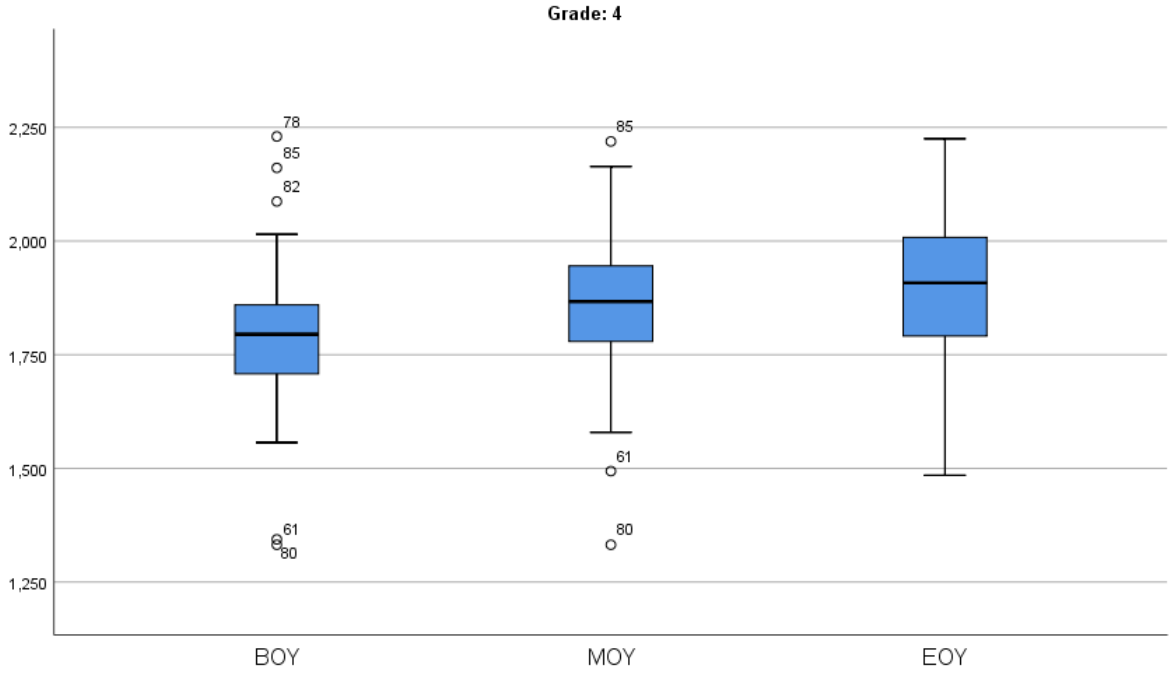

Figure 2. Boxplots for Grade 4.

As seen in Figure 2, none of the outliers were extreme points because, according to Laerd Statistics (2019), extreme points are more than three box-lengths away from the edge of their box. Since these outliers were not extreme, I chose to keep the outliers in

the analyses. I calculated a one-way repeated measures ANOVA with and without the outliers, compared the results, and found that the outliers had no effect on the analysis. The results were essentially the same because there was no statistically significant difference. Once these identified outliers were addressed, I moved on to Assumption 4 to determine the normality of the data.

RQ3 Outliers: Fifth grade ISIP scores. When analyzing the 55 data sets of fifth grade students' ISIP scores, I used the Explore: Plots tab to create boxplots for the data set. In the data set for Grade 5 (n = 67), these data rendered outliers for students 145 and 149. Figure 3 includes the boxplots and evaluated outliers. The reexamination of these data found no data entry or measurement errors. Therefore, they include genuinely unusual values. Further inspection and resolution of these outliers was the next step.

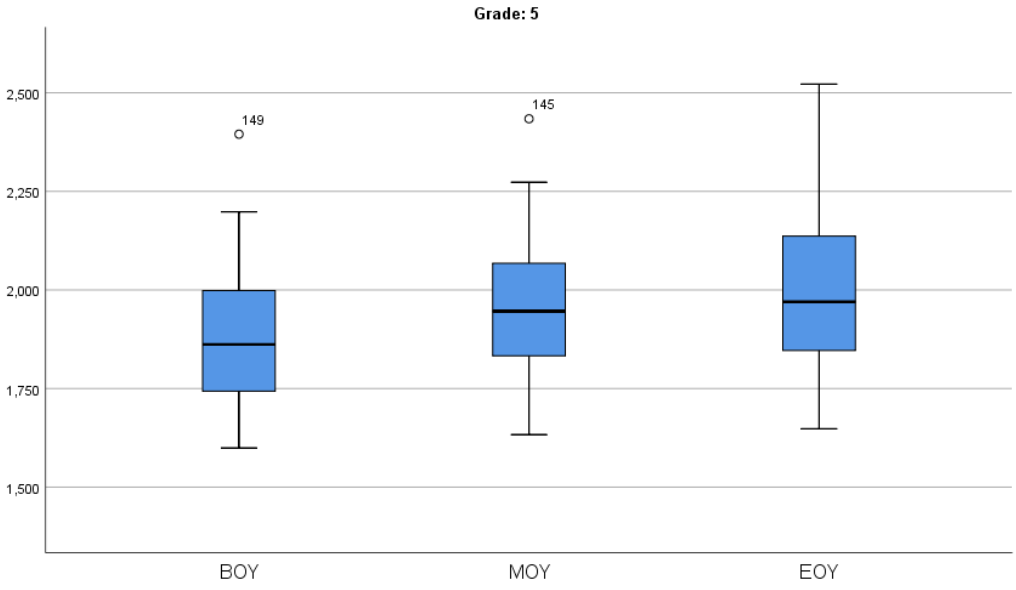

Figure 3. Boxplots for Grade 5.

As seen in Figure 3, none of the outliers were extreme points because, according to Laerd Statistics (2019), extreme points are more than three box-lengths away from the

edge of their box. Since these outliers were not extreme, I chose to keep the outliers in the analyses. I calculated a one-way ANOVA with and without the outliers, compared the results, and found that the outliers had no effect on the analysis. The results were essentially the same (i.e., no statistically significant difference). Once these identified outliers were addressed, I moved on to Assumption 4 to determine the normality of the data.

Assumption 4. Normality of the dependent variable is necessary to establish statistical significance (Laerd Statistics, 2019). The Grades 3-5 data sets included 53, 55, and 67 students, respectively, and I conducted a Shapiro-Wilk test for normality. In the Shapiro-Wilk test, if "data are normally distributed (i.e., the assumption of normality is met), the significance level . . . should be more than .05 (i.e., $p > .05$)" (Laerd, 2019, "Shapiro-Wilk Test of Normality," para. 2). The results for normality were different depending on grade level, and the BOY, MOY, and EOY for each of those grade levels and will be reported under each research question.

RQ1 Normality: Third grade ISIP scores. The normality assumption was not violated for time points BOY and MOY for Grade 3. For the BOY scores, Grade 3 results from the Shapiro-Wilk test reported $p = .18$ and $p = .52$ for MOY, which were both more than the required significance level of $p > .05$. However, though close, EOY rendered a violation of the normality assumption with $p = .044$, which is less than the required significance level of $p > .05$.

RQ2 Normality: Fourth grade ISIP scores. The normality assumption was violated for EOY in Grade 4 because the Shapiro-Wilk test resulted in $p = .023$.

According to Laerd Statistics (2019), this result is less than the required significance level of $p > .05$. The other two time points, MOY and EOY, were not violated with respective results of $p = .38$ and $p = .78$, which were both more than the required significance level of $p > .05$. Therefore, the EOY required further investigation regarding normality.

There are two options for dealing with violations regarding normality of the data. These include either (a) transforming the dependent variable or (b) carrying on regardless (Laerd Statistics, 2019). It is not unusual to see violations of this assumption with real-world data because the one-way repeated measures ANOVA is robust to violations of normality (Leech, Barrett, Morgan, Clay, & Quick, 2005). Therefore, I chose to carry on with these data for Grade 4 and move to testing for the final assumption.

RQ3 Normality: Fifth grade ISIP scores. When analyzing the 67 data sets for Grade 5, I ran the Shapiro-Wilk test again. Running the assumption for Grade 5 rendered results that met significance because in each level the significance was $p > .05$. The BOY score was $p = .12$, MOY was $p = .13$, and EOY was $p = .13$. Therefore, the assumption for normality was met for all the time points for Grade 5.

Assumption 5. Finally, Assumption 5 relates to equal variance of the dependent variable, known as sphericity. According to Tabachnick and Fidell (2017), *sphericity* is defined as measuring the correlation scores between the dependent variables that occur over time to determine whether they are similar. A violation of sphericity can lead to invalid results (Laerd Statistics, 2019). Since correlations are more likely to be similar between variables that are measured closer together in time, Mauchly's test can determine sphericity. Similarities between variables can lead to a Type I error, which

may show a statistically significant result when there is not one (Laerd Statistics, 2019). Using Mauchly's test statistic, sphericity can be either significant, meaning little probability of an error ($p < .05$) or nonsignificant ($p > .05$). If there is a violation of sphericity, additional statistical tests can be used to correct for the error (Tabachnick & Fidell, 2007).

RQ1 Sphericity: Third grade ISIP scores. For Grade 3, Mauchly's test of sphericity indicated that the assumption of sphericity had been violated, $\chi 2(2) = 17.9$, $p < .001$. In practice, the assumption of sphericity is considered difficult not to violate (Weinfurt, 2000). Further correction to account for this Type I error was necessary via either the Greenhouse-Geisser or Huynh-Feldt adjustment. According to Laerd Statistics (2019), when $\varepsilon > 0.75$, it is best to use the Huynh-Feldt correction. Furthermore, Abdi and Williams (2010) described this correction as more efficient and more powerful than the Greenhouse-Geisser correction. Epsilon (ε) was 0.791, as calculated according to the Huynh-Feldt adjustment, and was used to correct this one-way repeated measures ANOVA. The results were interpreted using this Huynh-Feldt adjustment.

RQ2 Sphericity: Fourth grade ISIP scores. Upon running the one-way repeated measures ANOVA for Grade 4, Mauchly's test of sphericity indicated that the assumption of sphericity had not been violated, $\chi^2(2) = 5.32$, $p = .07$. This means the test was not statistically significant and the assumption of sphericity was met. I can interpret these data based on the sphericity assumption and no further adjustment is necessary for Grade 4.

RQ3 Sphericity: Fifth grade ISIP scores. Upon running the one-way repeated measures ANOVA for Grade 5, Mauchly's test of sphericity indicated that the assumption of sphericity was not violated, $\chi^2(2) = 4.75$, $p = .09$. Therefore, the sphericity assumption was met and there is no need to account for a Type I error. I can interpret these data based on the sphericity assumption and no further adjustment is necessary for Grade 5.

Statistical Findings for the One-Way Repeated Measures ANOVA

Based upon the assumptions for the one-way repeated measures ANOVA and data sets for the first three research questions in this study, I calculated the one-way repeated measures ANOVA to determine whether there was a change in ISIP scores over time. I analyzed these data with descriptive statistics to determine the mean of each group's ISIP score at each point in time from the beginning to the end of the year. A summary of the analyses and respective results are listed under each research question.

RQ1 Results. A one-way repeated measures ANOVA was conducted to determine whether there were statistically significant differences in ISIP score over time throughout the course of the school year for Grade 3. The assumption of sphericity was violated, as assessed by Mauchly's test of sphericity, $\chi^2(2) = 17.9$, $p < .001$. Therefore, a Huynh-Feldt correction was applied ($\varepsilon = 0.791$). As seen in Table 3, the Istation program elicited statistically significant changes in scores over time, $F(1.581, 82.227) = 45.00$, $p < .005$, partial $\eta^2 = .46$, with ISIP scores increasing from both the initial assessment (BOY3: $M = 138.09$, $SD = 14.80$) to mid-year (MOY3: $M = 249.28$, $SD = 15.64$) and the initial assessment to end of year (EOY3: $M = 250.28$, $SD = 20.62$), but not from MOY3

to EOY3. Therefore, the null hypothesis is rejected. Table 3 includes the Grade 3 statistics for the one-way repeated measures ANOVA.

Table 3

Grade 3 Mean, Standard Deviation, and Number of Student ISIP Scores (BOY, MOY, EOY) from Istation Reading Program

ISIP	*M*	*SD*	*N*
BOY3	238.09	14.81	53
MOY3	249.28	15.65	53
EOY3	250.28	20.62	53

Note. BOY = beginning of year; MOY = middle of year; EOY = end of year

Post hoc analysis with a Bonferroni adjustment revealed the ISIP score was statistically significantly increased from BOY3 to MOY3 and from BOY3 to EOY3 (95% CI [8.67, 13.71], $p < .001$). The Bonferroni adjustment also indicated no statistical significance for MOY3 to EOY3 ($M = 250.28$, 95% CI [2.60, 4.60], $p = 1.00$). This finding indicates that the difference between times was significant only from BOY3 to MOY3 and BOY3 to EOY3, not from midpoint to end of year.

RQ2 Results. A one-way repeated measures ANOVA was conducted to determine whether there were statistically significant differences in ISIP score over time throughout the course of the school year for Grade 4. As seen in Table 4, the Istation program elicited statistically significant changes over time, $F(2, 108) = 58.14$, $p < .005$, partial $\eta^2 = .52$, with ISIP scores increasing from both the initial assessment (BOY4: $M =$ 1,785.96, $SD = 158.74$) to mid-year (MOY4: $M = 1,863.60$, $SD = 170.96$), the initial

assessment to end of year (EOY4: *M* = 1,899.89, *SD* = 158.00), and the mid-year to end of year. Therefore, the null hypothesis is rejected.

Table 4

Grade 4 Mean, Standard Deviation, and Number of Student ISIP Scores (BOY, MOY, EOY) from Istation Reading Program

ISIP	*M*	*SD*	*N*
BOY4	1,785.96	158.74	55
MOY4	1,863.60	170.96	55
EOY4	1,899.89	158.00	55

Post hoc analysis with a Bonferroni adjustment revealed the ISIP score was statistically significantly increased from BOY4 to MOY4, from BOY4 to EOY4 (95% CI [48.78, 106.50], $p < .001$), and from MOY4 to EOY4 (*M* = 1,899.89, 95% CI [14.11, 58.47], $p = .001$). These findings indicate that student reading improved with use of Istation from the beginning to middle of the year and from the beginning to the end of the year. Additionally, in Grade 4 only, there was a significant change from the middle to end of year. Therefore, a change in reading from the middle to end of the year can be statistically attributed to Istation use.

RQ3 Results. A one-way repeated measures ANOVA was conducted to determine whether there were statistically significant differences in ISIP score over time throughout the course of the school year for Grade 5. As seen in Table 5, the Istation program elicited statistically significant changes over time, $F(2, 130) = 35.85$, $p < .005$, partial $\eta^2 = .36$, with ISIP scores increasing from both the initial assessment (BOY5: *M* =

1,893.88, *SD* = 169.64) to mid-year (MOY5: *M* = 1,953.80, *SD* = 169.63) and the initial assessment to end of year (EOY5: *M* = 1,990.15, *SD* = 195.70). Therefore, the null hypothesis is rejected.

Table 5

Grade 5 Mean, Standard Deviation, and Number of Student ISIP Scores (BOY, MOY, EOY) from Istation Reading Program

ISIP	*M*	*SD*	*N*
BOY5	1,893.88	169.64	67
MOY5	1,953.80	169.63	67
EOY5	1,990.15	195.70	67

Post hoc analysis with a Bonferroni adjustment revealed the ISIP score was statistically significantly increased from BOY5 to MOY5, from BOY5 to EOY5 (95% CI [33.90, 85.95], $p < .001$), but not from MOY5 to EOY5 (*M* = 1,990.15, 95% CI [4.59, 68.11], $p = .02$). These findings indicate that student reading improved with use of Istation from the beginning to middle of the year and from the beginning to the end of the year. However, any change in reading from the middle to end of the year cannot be statistically attributed to Istation use.

Summary. Overall, results of the ANOVA analyses show that there is a change in ISIP scores over time. For Grades 3 and 5, the change over time was statistically significant from BOY to MOY and from BOY to EOY but not from MOY to EOY. However, the Grade 4 change over time was statistically significant for all time points

(BOY → MOY, BOY → EOY, and MOY → EOY). This finding is an indication that the students are learning with the use of the Istation reading formative assessment program.

Linear Regression Statistical Assumptions

The intent of a linear regression is to analyze the extent of the predictive ability of an independent variable on a dependent variable. According to Tabachnik and Fidell (2007), "The flexibility of techniques is, then, especially useful to the researcher who is interested in real-world or very complicated problems that cannot be meaningfully reduced to orthogonal designs in a laboratory setting" (p. 111). It is effective with continuous data and strives to create regression coefficients. In this study, a simple linear regression was performed to create an equation that would best quantify the relationship between the independent variable, ISIP MOY reading formative assessment score, and the ELA PARCC summative assessment score dependent variable.

For the linear regression to be an appropriate analysis, seven assumptions about the data must be met. According to Laerd Statistics (2019), verifying that a data set meets these assumptions is key to interpreting the validity of the results. However, it is common in real-world data for a data set to fail an assumption (Laerd Statistics, 2019). In these instances, it is essential to apply appropriate solutions and possibly further testing to overcome the violation of the assumption. The seven assumptions for a linear regression are as follows:

1. The data must contain a continuous dependent variable.
2. The data must have a continuous independent variable.
3. There is a linear relationship between the two variables.

4. There must be independence of observations.
5. There are no significant outliers.
6. There is homoscedasticity.
7. The residuals (errors) of the regression line are approximately normally distributed. (Laerd Statistics, 2019)

Assumptions 1–2. According to the first assumption, the dependent variable in this study must contain continuous data (Laerd Statistics, 2019). In this study, each student's ELA PARCC raw score represents the dependent variable (with a range from 650 to 850). The second assumption involves the requirement of a continuous independent variable. Based on this data set, the Istation formative assessment MOY score is the independent variable, which is tabulated on a continuous scale (i.e., values for Grade 3 range from 196–293; Grade 4, 1,330–2200; Grade 5, 1,600–2,600). As both the variables were tabulated on a continuous scale, these data met the first two assumptions. The next three assumptions involved the nature of the data.

Assumption 3. The third assumption involves the need for a linear relationship between the independent and dependent variable. According to Laerd Statistics (2019), the way to assess a linear relationship is to inspect a scatter plot. If the relationship approximately follows a straight line, there is a linear relationship. However, if there is, for example, a curved line, there is no linear relationship (Laerd Statistics, 2019). I used the Explore tab in SPSS to create scatter plots and then evaluated them for each set of grade-level data to verify a linear relationship between the two variables.

RQ4 Linear Relationship: Third grade ISIP and ELA PARCC scores. Scatter plots for Grade 3 ELA PARCC raw scores compared to the MOY ISIP scores from the Istation formative assessment program were plotted. As indicated in Figure 4, visual inspection of these data indicated a linear relationship between the variables as the data approximately resembled a straight line. This assumption was met; and it was appropriate to move on to the fourth assumption: independence of residuals.

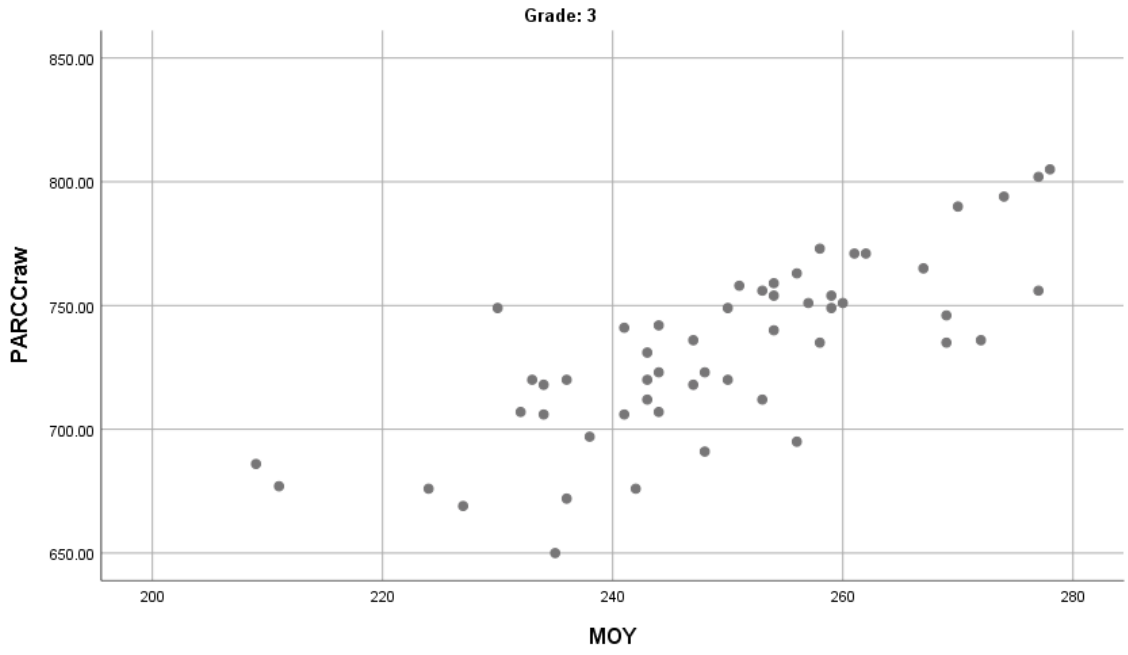

Figure 4. Scatter plot for Grade 3.

RQ5 Linear Relationship: Fourth grade ISIP and ELA PARCC scores. Scatter plots comparing Grade 4 ELA PARCC raw scores to the MOY ISIP scores from the Istation formative assessment program were plotted. As seen in Figure 5, visual inspection of these data indicated a linear relationship between the variables via an approximately straight line. This assumption was met, and it was appropriate to move on to the fourth assumption.

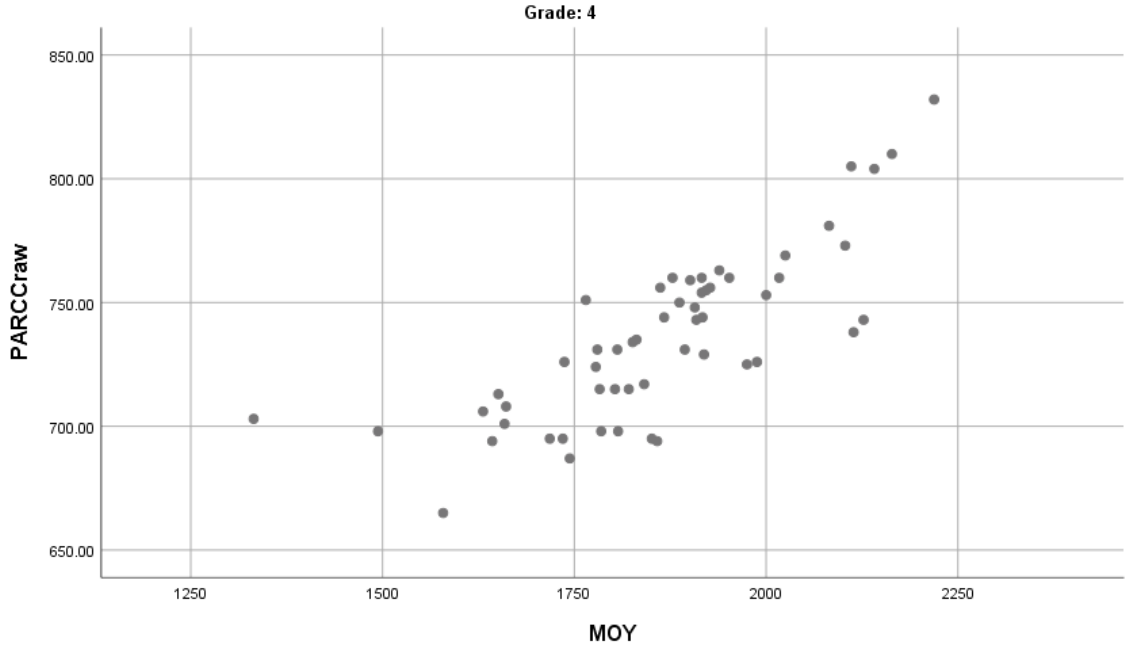

Figure 5. Scatter plot for Grade 4.

RQ6 Linear Relationship: Fifth grade ISIP and ELA PARCC scores. I used a scatter plot to evaluate Grade 5 ELA PARCC raw scores compared to the MOY ISIP scores from the Istation formative assessment program. As seen in Figure 6, visual inspection of these data indicated an approximately straight line and, therefore, a linear relationship between the variables. This assumption was met, and it was appropriate to move on to the fourth assumption: independence of residuals.

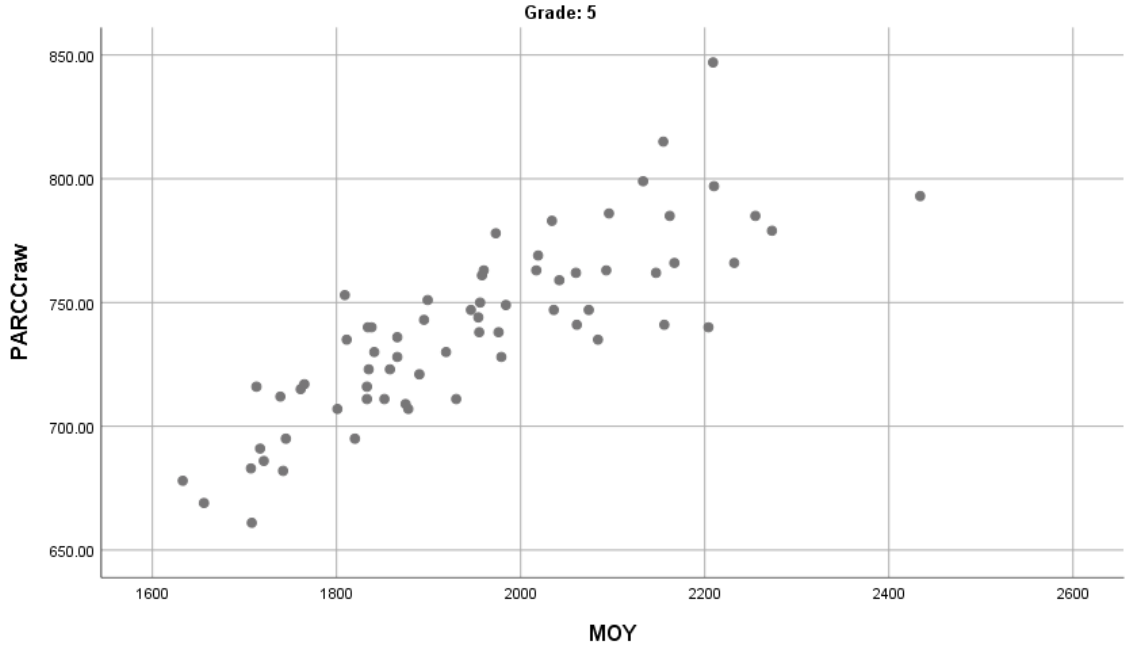

Figure 6. Scatter plot for Grade 5.

Assumption 4. After evaluation of Assumptions 1–3, it was necessary to run the linear regression analysis in SPSS to address the final four assumptions. The fourth assumption involved ensuring the data do not overlap or offer information about each other, which is known as independence of observations (Laerd Statistics, 2019). Each observation must stand alone and must not be related. Independence of observations can be assessed by evaluating the Durbin-Watson test. According to Laerd Statistics (2019), the Durbin-Watson statistic ranges from 0 to 4, and to evaluate these data to see whether there is independence of observations, "it is crucial to find a value of approximately 2, which indicates that there is no correlation between residuals" ("Assumptions II," para. 5). If the Durbin-Watson results reveal a statistic of approximately 2, this assumption is considered met as each piece of these data appears to stand alone.

RQ4 Independence of Observation: Third grade ISIP and ELA PARCC scores. When analyzing the third grade data sets, I tested for the assumption of independence of observations by using the Durbin-Watson tests. Findings revealed a statistic of 2.20, which meets the requirement of being close to 2. Therefore, there was independence of residuals for the data sets of Grade 3.

RQ5 Independence of Observation: Fourth grade ISIP and ELA PARCC scores. The independence of residuals assumption for Grade 4 was also met. These data, as assessed by the Durbin-Watson test, rendered a result of 2.21, which meets the requirement of being close to 2.

RQ6 Independence of Observation: Fifth grade ISIP and ELA PARCC scores. There was independence of residuals for the data sets of Grade 5, as assessed by the Durbin-Watson statistic of 2.20. This result meets the requirement of being close to 2 as well as the assumption. Therefore, it was appropriate to move on to the next assumption.

Assumption 5. The fifth assumption involves the presence and potential mitigation of outliers. Outliers, or unusual points, can be present in data and must be evaluated to ensure accuracy of the regression line (Laerd Statistics, 2019). Checking this assumption involves diagnosing the regression model for cases in the data that are mitigating outliers that exert significant influence on the model. To accomplish this task, a Cook's distance analysis was conducted (Leech et al., 2015). With results over 1, it is necessary to further investigate and potentially remove the outliers. When there are outliers in the data set—as there were in this study—best practice requires an examination of the data set for (a) data entry errors, (b) measurement errors, or (c)

genuinely unusual values (Laerd Statistics, 2019). However, if the Cook's distance analysis results are less than 1 for all these data, there are no outliers, and it is appropriate to move on to Assumption 6.

RQ4 Outliers: Third grade ISIP and ELA PARCC scores. To address Assumption 5, Cook's distance was applied to the analysis. For Grade 3, Cook's distance results indicated a minimum of .00 and maximum of .13. This means there were no outliers as neither number is over 1. Hence, this assumption was met.

RQ5 Outliers: Fourth grade ISIP and ELA PARCC scores. The fifth assumption involves mitigating outliers. For Grade 4, the results indicated a minimum of .00 and a maximum of .91. Though not over 1, the maximum is close to 1, which required further investigation. To address the influential outlier in the Grade 4 data, it is necessary to review the Cook's distance variable in SPSS and see which score is the outlier. Upon examination of these data, there were no entry or measurement errors. Therefore, the outliers were determined to be genuinely unusual values. With unusual values, best practice involves determining whether the outlier should be kept or removed from the data set.

To evaluate whether to keep or remove the influential outlier from analysis, I removed the outlier (i.e., Case 81) and ran the regression analysis again. The Cook's distance results revealed no other significant influential cases for Grade 4 (minimum of .00 and maximum of .21), and the elimination of this case also improved normality, the relationship among the residuals, and heteroscedasticity as revealed by the residual plot. This visual inspection was justification to exclude Case 81. This plan ensured that the one

specific data point did not overly influence the rest of the data set. With this case removed, the rest of the results will use this new data set (n = 54), and it was appropriate to move to the sixth assumption.

RQ6 Outliers: Fifth grade ISIP and ELA PARCC scores. To address Assumption 5, Cook's distance was applied to the analysis. For Grade 5, the results showed a minimum of .00 and a maximum of .26, indicating there were no excessively influential outliers in these data. It was appropriate to move on to Assumption 6.

Assumption 6. This assumption regards homoscedasticity. This assumption tests to see whether the residuals are equal across the predicted values (Laerd Statistics, 2019). Evaluating this assumption can be checked via inspection of a scatter plot of the residuals. If examination of these data indicate heteroscedasticity, there are ways to counteract the patterns in the residuals (Laerd Statistics, 2019). However, if the residuals are equal across the predicted values, there is homoscedasticity (Laerd Statistics, 2019). This means there is a constant spread and no pattern in these data points, the assumption is met, and it is appropriate to move on to assess the next assumption.

RQ4 Homoscedasticity: Third grade ISIP and ELA PARCC scores. The third grade data's homoscedasticity was assessed with the use of a scatter plot. There was homoscedasticity as assessed by visual inspection of a plot of the Grade 3 standardized residuals versus standardized predicted values. There was a constant spread of the residuals and no pattern existed. Therefore, this assumption was met and it was appropriate to move to the seventh assumption.

RQ5 Homoscedasticity: Fourth grade ISIP and ELA PARCC scores. The fourth grade data's homoscedasticity was assessed with the use of a scatter plot. There was homoscedasticity as assessed by visual inspection of a plot of Grade 4 standardized residuals versus standardized predicted values. There was a constant spread of the residuals and no pattern existed. Therefore, this assumption was met and it was appropriate to move to the seventh assumption.

RQ6 Homoscedasticity: Fifth grade ISIP and ELA PARCC scores. The fifth grade data's homoscedasticity was assessed with the use of a scatter plot. There was homoscedasticity as assessed by visual inspection of a plot of standardized residuals versus standardized predicted values. There was a constant spread of the residuals and no pattern existed. Therefore, this assumption was met and it was appropriate to move to the seventh assumption.

Assumption 7. The seventh and final assumption involves checking for normality of residuals. Two methods, a histogram or a normal probability (P-P) plot, can be used to assess residual normality. These plots are generated automatically if selected in the plots box when running the linear regression analysis in SPSS. Laerd Statistics (2019) stated that a normal P-P plot is one of the best graphical methods to assess normality. If residuals are normally distributed, data points will be approximately aligned along the diagonal line. However, Laerd Statistics stated that if the data points are not approximately aligned along the diagonal line of the P-P plot, it is necessary to "transform the dependent variable to try to coax the error residuals to normality"

("Checking for Normality of Residuals," para. 8). The specific grade-level data's normality as assessed by the normal P-P plot is described under each research question.

RQ4 Normality: Third grade ISIP and ELA PARCC scores. The seventh and final assumption regards the normal distribution of residuals. For Grade 3, residuals were normally distributed as assessed by visual inspection of a normal P-P plot. After these seven assumptions were met, it was appropriate to further interpret the results.

RQ5 Normality: Fourth grade ISIP and ELA PARCC scores. The seventh and final assumption regards the normal distribution of residuals. For Grade 4, residuals were normally distributed as assessed by visual inspection of a normal P-P plot. After these seven assumptions were met, it was appropriate to further interpret the results.

RQ6 Normality: Fifth grade ISIP and ELA PARCC scores. The seventh and final assumption regards the normal distribution of residuals. For Grade 5, residuals were normally distributed as assessed by visual inspection of a normal P-P plot. After these seven assumptions were met, it was appropriate to further interpret the results.

Statistical Analysis Findings for Linear Regression

To address the final three research questions regarding the relationship between the Istation formative assessment program and its ability to influence the ELA PARCC scores, I ran a linear regression analysis. I analyzed these data to identify the percentage of variance, evaluate the statistical significance of the model, and interpret the coefficients. Finally, I created a regression equation.

There are important measures to interpret and report in a linear regression model. I performed a quantitative analysis using SPSS Version 25.0, a standard confidence

interval of 95%, and an alpha of 0.05. The conventional medium effect size of 0.25 was used (Cohen, 1992). First, it is necessary to interpret the adjusted R^2 because it represents the proportion of variance that is explained by the independent variable. The adjusted R^2 is used over the R figure because, "it corrects for the positive bias in order to provide a value that you would expect in the population" (Laerd Statistics, 2019, p. 16). I interpreted the results using the output tables titled Model Summary and ANOVA. The analysis requires evaluation of the effect size via adjusted R^2, examination of whether or not the coefficients show a linear relationship between the two variables, and evaluation of the ability of the independent variable (i.e., Istation program/ISIP score) to predict the dependent variable (i.e., ELA PARCC score).

RQ4 Results. A linear regression was conducted to determine whether the Istation formative assessment MOY ISIP Grade 3 student score could predict performance on the ELA PARCC summative assessment. The analysis established that, for Grade 3, a student's ISIP score from the MOY Istation formative assessment statistically significantly predicts the student's score on the ELA PARCC summative assessment ($F[1, 51] = 73.54$, $p < .001$), and the ISIP score accounted for 58.2% of the explained variability in the ELA PARCC score as assessed using the adjusted R^2 measure. Therefore, the null hypothesis is rejected. This means that 41.8% of the variance in the ELA PARCC is explained by other factors that were not measured in this analysis.

I completed a linear regression to investigate the relationship between the independent and the dependent variables. Results rendered an equation that represents the

relationship between the independent variable (MOY ISIP sore) and the dependent variable (ELA PARCC). Table 6 includes the data used to calculate this equation and the regression analysis summary, means, and standard deviations for Grade 3 data. The regression equation for MOY that predicts ELA PARCC (Y) was as follows: $Y = bX + a$. The value for *a* can be found in the B column and is 300.81. Therefore, for Grade 3, the predicted ELA PARCC = 1.7(MOY) + 300.81.

Table 6

*Regression Analysis Summary, Means, and Standard Deviations for Reading Achievement PARCC and ISIP MOY Scores for Grade 3 (*n *= 53)*

Variable	*M*	*SD*	*MOY*	*B*	*SEB*	β
PARCC score	731.38	35.17	.77*	300.81	50.31	
Independent variable						
ISIP MOY	249.28	15.65		1.73	.20	.77

Note. Adjusted $R^2 = .58$, $F(1,51) = 73.54$
*$p < .001$

RQ5 Results. I conducted a linear regression to determine whether the Istation formative assessment MOY ISIP Grade 4 student score could predict performance on the ELA PARCC summative assessment. This analysis established that, for Grade 4, a student's ISIP score from the MOY Istation formative assessment could statistically significantly predict the student's score on the ELA PARCC summative assessment (F[1, 52] = 108.00, $p < .001$), and the ISIP score accounted for 67% of the variance in the ELA PARCC score as assessed using the adjusted R^2 measure. Therefore, the null hypothesis

is rejected as the MOY ISIP score can predict the ELA PARCC score. The regression equation was: predicted ELA PARCC = .18(MOY) + 402.60. Table 7 includes the regression analysis summary, means, and standard deviations for Grade 4 data.

Table 7

*Regression Analysis Summary, Means, and Standard Deviations for Reading Achievement PARCC and ISIP MOY Scores for Grade 4 (*n = *54)*

Variable	*M*	*SD*	*MOY*	*B*	*SEB*	β
PARCC score	735.91	33.84	.79*	443.28	30.98	735.91
Independent variable						
ISIP MOY	1,863.60	170.96		.16	.02	.79

Note. Adjusted $R^2 = .62$, $F(1,52) = 89.98$
*$p < .001$

RQ6 Results. I conducted a linear regression to determine whether the Istation formative assessment MOY ISIP Grade 5 student score could influence performance on the ELA PARCC summative assessment. A linear regression established that, for Grade 5, a student's ISIP score from the MOY Istation formative assessment could statistically significantly predict the student's score on the ELA PARCC summative assessment ($F[1, 65] = 146.34$, $p < .001$), and the ISIP score accounted for 68.8% of the variance in the ELA PARCC score as assessed using the adjusted R^2 measure. Therefore, the null hypothesis is rejected as the MOY ISIP score can predict the ELA PARCC score. The regression equation was: predicted ELA PARCC = .17(MOY) + 402.57. Table 8 includes the regression analysis summary, means, and standard deviations for Grade 5 data.

Table 8

*Regression Analysis Summary, Means, and Standard Deviations for Reading Achievement PARCC and ISIP MOY Scores for Grade 5 (*n = *67)*

Variable	*M*	*SD*	*MOY*	*B*	*SEB*	β
PARCC score	739.57	35.92	.83*	402.57	27.97	
Independent variable						
ISIP MOY	1,949.01	172.84		.17	.01	.83

Note. Adjusted $R^2 = .69$, $F(1,65) = 146.34$
*$p < .001$

Summary of linear regression. The linear regression results addressed the second set of research questions and provided evidence that a single predictor model that includes Istation performance at MOY (i.e., Time 2) can be used to successfully predict student achievement on the ELA PARCC. This finding is a strong indication that participation in the Istation program may exert a significant influence on students' literacy performance as measured by the ELA PARCC test.

Summary of Research Questions

All six research questions were answered to find null hypotheses rejected and alternate hypotheses accepted. The Istation formative assessment program elicited a change over time for all grades from BOY to EOY and from MOY to EOY for Grade 4. Also, the Istation formative assessment was found to be predictive in its ability to influence the ELA PARCC summative assessment. These results support both the connection between formative and summative assessment and the ability of the Istation

formative assessment score to predict the ELA PARCC score, which will be explained in detail in Chapter 5.

Summary

In this chapter, I outlined the data collection, statistical analyses, and findings to the research questions for this study. The study investigated the change in ISIP score over time and the overall influence of the reading Istation computer-based formative assessment program on summative student achievement on the ELA PARCC. The one-way repeated measures ANOVA and linear regression analyses were used to address the research questions. Results revealed statistically significant changes over time for all time points for Grade 4 including BOY to MOY, MOY to EOY, and BOY to EOY. There was also statistical significance over time for BOY to MOY and BOY to EOY for Grades 3 and 5. The linear regression revealed the predictive ability of the MOY on ELA PARCC. Chapter 5 includes a more detailed report of the findings and recommendations of the study.

Chapter 5: Discussion, Conclusions, and Recommendations

The nature of this study was to determine to what extent the influence of the Istation individualized, computer-based, formative assessment program in reading had on student achievement in Grades 3–5. The purpose of this quantitative study was to investigate the influence of the Istation reading formative assessment program on student achievement as measured by the ELA PARCC summative assessment. The study was conducted to address the gap in practice regarding the link between formative and summative assessment and to add to the quantitative research in the field.

To respond to the first three research questions, I used a one-way repeated measures ANOVA. The dependent variable was the Istation formative assessment program. The independent variable was each student's individual ISIP scores from the three points in time throughout the school year the students took the Istation assessment: BOY, MOY, and EOY. Key findings included the significant difference in the mean ISIP scores and statistically significant changes over time from BOY to MOY, BOY to EOY, and MOY to EOY for Grade 4. There was also statistical significance over time from BOY to MOY and from BOY to EOY for both Grades 3 and 5.

To respond to the second set of research questions, I used a linear regression analysis. The dependent variable was the ELA PARCC score while the independent variable was the Istation formative assessment (MOY ISIP score). Key findings included the statistically significant ability of the Istation formative assessment program to influence the ELA PARCC scores. In this chapter, I present an interpretation of the results situated within the context of the literature review and the theoretical foundation

for this study. Also included is a discussion regarding the limitations of the study, recommendations, and implications for future research and practice.

Interpretation of the Findings

The current study examined the Istation formative assessment program and its relationship to summative assessment—specifically, to the ELA PARCC. Teachers and administrators historically use formative assessment practices to understand where students are and where they need to be on their path of academic achievement and skill acquisition. The Istation program's designers assert that the program individualizes student feedback for faster skill acquisition. In this section, I provide interpretations based on results of both analyses and then compared these interpretations with the research found in the literature review. I also include the connection to Sadler's theoretical framework.

Comparison to the Literature

The affirmative answers to the research questions confirmed and aligned with research discussed in Chapter 2, supporting formative assessment practices in general (De Lisle, 2015; Herman et al., 2015; Karim, 2015). Likewise, the results from this study confirmed research that the Istation formative assessment program may influence and predict student achievement (Luo et al., 2017; Marin, 2015; Patarapichayatham, 2014). This section includes a comparison of the overarching themes in the previous literature about the link between formative and summative assessment and specifically the Istation program's ability to influence the ELA PARCC assessment scores and ultimately improve student achievement.

RQ 1–3: Formative Assessment and Literacy. This study aligns with researchers who found an improvement in literacy skills such as the Istation program assesses. These skills included grammar, textual, functional, and sociolinguistic language (see Barefoot, 2017; Boumediene & Hamzaoui-Elachachi, 2017; Bulat et al., 2017). Bennett et al. (2017) found similar positive results with text comprehension and fluency. Previous research conducted by Barefoot (2017) revealed specific gains in motivation to conduct research when writing, which is also a component of the Istation reading program. The one-way repeated measures ANOVA results indicated a statistically significant increase in student scores over time for all grades, which was evident from the beginning to the middle of the year and the beginning to the end of the year. For all grades, there was a statistically significant increase in mean scores from the beginning to the middle of the year and the beginning to the end of the year. However, there was only a change from the middle to the end of the year in Grade 4.

These findings from the current study align with the recent studies mentioned above because the formative assessment helps close the gap in knowledge from where the students are to where they should be (Barefoot, 2017; Boumediene & Hamzaoui-Elachachi, 2017; Bulat et al., 2017). The beginning score was a baseline, and use of the formative assessment Istation improved student scores from the baseline to the midpoint and certainly from beginning to end of year. However, there was no previous literature addressing instances such as this one, where the increase from MOY to EOY was insignificant, except in Grade 4. This finding could indicate the need for future research in Istation implementation processes or mitigating factors that may have affected student

use of or output from Istation. This research may help teachers understand how students learn over time and throughout the year.

Computer-Based Formative Assessment. The findings from studies on computer-based formative assessment are aligned to this study as they both generated positive results. Meta-analyses of various computer-based formative assessment learning tools found them to be useful to track student achievement and measure progress (see Belo et al., 2016; Shute & Rahimi, 2017). Studies completed in Sweden, China, and Australia had similar findings, and researchers suggested a significant potential use of digital formative assessments (see Bhagat & Spector, 2017; Cloonan et al., 2016; Genlott & Grönlund, 2016). The results from this study indicated the Istation computer-based formative assessment student scores improve over time and that the tool may predict student performance on the summative assessment. These findings reinforce the knowledge from previous literature that computer-based assessment has a crucial role in elementary education.

RQ 4–6: Formative and Summative Assessment and Student Achievement. There has been an overarching agreement among researchers and educators that formative assessment is a helpful tool for student achievement. Meta-analyses conducted between 2014 and 2019 found formative assessment to be the preferred method to evaluate progress and found summative assessment to be a snapshot of learning (see Buelin et al., 2019; Lau, 2016). The link between the two types of assessment continues to be investigated in studies concerning specific programs and a variety of content areas.

Several researchers found that formative assessment practices prepared students for summative assessments. According to Aydin and Ürün (2016), scores from pre- to posttest improved with the use of formative assessment. Huang (2016) reported an improvement in student achievement and memorization with the use of formative assessment. The findings of a study by Ozan and Kincal (2018) indicated similar results, with students improving academically and having better attitudes toward learning. Simmons et al. (2015) found a steady progression of the curriculum for those students who took the formative tests versus the control group. The findings of this study revealed a significant ability of the Istation reading program to influence and predict student achievement on the ELA PARCC—thus confirming the argument for using formative assessment to improve student achievement.

Istation. Though no research existed regarding the Istation formative assessment program's influence on summative achievement or specifically the ELA PARCC, there were studies that involved other similar ELA or reading skill assessment. The study that analyzed data from 98 third grade children by Luo et al. (2017) revealed a strong correlation between the Istation program and the STAR reading assessment. Other studies revealed the ability of the Istation program to predict student achievement (Marin, 2015; Patarapichayatham, 2014). Due to the lack of prior research involving the PARCC, the findings of this study were the first of their kind in examining the Istation formative assessment program and its influence on the ELA PARCC summative assessment.

Findings revealed that mean scores on the Istation formative assessment program changed over time. Also, results indicated that the Istation program scores can predict

student performance on the ELA PARCC. A regression equation predicted ELA PARCC = 300.81 + 1.7(MOY), showing the ability of formative assessment to influence summative assessment. These findings support knowledge of this program as a good predictor of other summative assessments like the STAR and STAAR. This finding is aligned with the previous literature on the Istation formative assessment program.

Findings and Sadler's Formative Assessment Framework

Sadler's formative assessment framework guided this study. The formative assessment cycle begins with assessing where students are regarding skill level, identifying where they need to be, and determining how to close the gap in knowledge to get them there (Sadler, 1989). The Istation's designers claim their program individualizes feedback to help learners get where they need to be. Reading skills are assessed monthly at HES, and the ability of the Istation program to complete the cycle of feedback outlined in Sadler's work was unknown. Therefore, investigating whether there is a change in score over time with the use of the Istation program was one of two major focuses of this study.

The second focus of this study was to see whether the Istation program completes the cycle of feedback by closing the gap in knowledge for students. In most elementary settings, the summative assessment score typically indicates student skill acquisition and achievement. Therefore, investigating the ability of the Istation formative assessment program (MOY score) to influence and predict summative achievement on the ELA PARCC was necessary to see whether this formative assessment can help teachers guide

instruction and improve student achievement. The findings are in alignment with Sadler's framework and previous studies citing Sadler.

As seen in Tables 3–5, the results of the one-way repeated measures ANOVA for Grades 3–5 suggest that Sadler's formative assessment cycle was completed because findings indicate a change over time in student scores using the Istation program. As seen in Tables 6–8, the findings from the linear regression analysis and RQ4–6 also indicate the effectiveness of the Istation formative assessment to predict student achievement. This ability offers teachers the chance to help students gain the skills they need to reach grade-level skills. Therefore, it is clear the program analyzed in this study can complete the cycle of Sadler's formative assessment framework.

Limitations of the Study

This study was limited to one school building that was part of one Title I school district in a rural southwest town. Though the sample represented a diverse population with low socioeconomic status, it only represented the students from that one school, town, and surrounding areas. Second, the study only examined the scores for Overall Reading, which was the general score the Istation program automatically assigned based on student performance. Therefore, the subsets of text fluency, comprehension, and spelling were not saved in the archives and not included. If these data were available, it would have given a better understanding of which areas the students improved in over time, and perhaps some of the gaps in knowledge that were not indicative of a statistically significant change. This study's results can be generalized to students in Grades 3–5 but not Grade 6 due to the issue with initial BOY scores for that group. Finally, since

teachers were not required to keep a log of minutes students interacted with the program each month, there was no way to tell whether students got the suggested time in with the program to see optimal results.

Recommendations

Future research that expands to other elementary settings in other geographic regions is recommended. Including larger samples from other schools would also be helpful. Future research could expand on sample size and student demographics, as well as specific literacy skills such as reading comprehension, text fluency, and spelling to see how Istation performs as a formative assessment in those specific areas. Since the current study was limited to Grades 3–5, it would be essential to learn more about how the Istation program is used to support learners in other grades.

Future research would be essential to learn more about how the Istation program is used in between each monthly assessment. This research could explore the time students spent on the program compared to their success in gaining grade-level skills over time. Future studies could include the frequency of the assessment and whether monthly testing was necessary or if BOY, MOY, and EOY suffice. This study was related to reading, but Istation also has a math component. Investigating the math student achievement could provide useful information about the efficacy of the program.

Finally, the current study did not include information about the training teachers and students received prior to or during administration of this assessment. Future research could include the difference in student performance based on the amount of teacher training involved before implementation. Also, it would be helpful to investigate the

difference in self-efficacy for teachers who buy-in to the program and those who do not. Finally, another worthy endeavor would be investigating the difference in scores for students who have received training for use of the program and those who did not to see whether there was an improvement in their scores over time.

Implications

The results from this study may promote positive social change in a way that can inform teachers, parents, school administrators, and policy makers about the effect of computer-based formative assessment to improve student literacy and achievement. Analyzing these findings allows people teaching in the elementary school setting to embrace digital assessments as a helpful tool. These assessments appear to accurately gather and analyze data at the individual student level, personalize instruction, and save the teacher precious planning time.

The expectations of the summative assessment design and public education is to graduate students that are college or career ready. The use of technology to compete in our society is essential. Teaching students to use digital tools for learning is one of the pieces to move toward a college- or career-ready population. Furthermore, with the state of education in 2020, Istation could be used as a helpful tool for elementary students to access during virtual learning. The feedback could be automatically sent to teachers to gain a better understanding of student skill levels. This study contributes to the education practice because there is now more research-based evidence that the Istation program is a useful tool for students and teachers in our ever-changing digital world.

Conclusion

In this study, I sought to determine the relationship between performance on the Istation reading formative assessment program and student achievement on the ELA PARCC summative assessment. A review of the literature indicated that, overall, there is agreement in the field of education about formative assessment being a useful tool for teachers (De Lisle, 2015; Herman et al., 2015; Karim, 2015; Owen, 2016). However, little research existed looking at specific formative assessments like Istation. I sought to fill the gap in practice regarding the link between formative and summative assessment. An archived data set comprised of three Istation reading programs and ELA PARCC scores were analyzed through a one-way repeated measures ANOVA and linear regression. The findings of these statistical analyses indicate that implementation of the Istation reading formative assessment program improved student scores over time and the program's MOY score can predict student outcomes on the ELA PARCC. The results of this study generated knowledge about whether Istation can be a useful assessment tool toward student achievement in Grades 3–5 and worth the financial investment for schools. Though future research is needed in these grade levels, this study begins the conversation about using Istation to help students on their path to grade-level skill acquisition and help teachers to move their students toward success on summative achievement.

References

Abdi, H., & Williams, L. J. (2010) Principal component analysis. *Wiley Interdisciplinary Reviews: Computational Statistics, 2*, 433–459. https://doi.org/10.1002/wics.101

Aljezawi, M., & Albashtawy, M. (2015). Quiz game teaching format versus didactic lectures. *British Journal of Nursing*, *24*(2), 86–92. https://doi.org/10.12968/bjon.2015.24.2.86

All, A., Nunez Castellar, E. P., & Van Looy, J. (2015). Towards a conceptual framework for assessing the effectiveness of digital game-based learning. *Computers & Education, 88*, 29–37. https://doi.org/10.1016/j.compedu.2015.10.007

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for Educational and Psychological Testing*, *1*, 99.

Andersson, C., & Palm, T. (2017). Characteristics of improved formative assessment practice. *Education Inquiry (Co-Action Publishing)*, *8*(2), 104–122. https://doi.org/10.1080/20004508.2016.1275185

Apthorp, H., Klute, M., Petrites, T., Harlacher, J., & Real, M. (2016). Valuing a more rigorous review of formative assessment's effectiveness. *Society for Research on Educational Effectiveness*, 1–9.

Aydin, S., & Ürün, N. (2016). The effect of formative assessment technique on academic success of the students and their attitudes in the unit "The Solar System and Beyond: The Space Puzzle" at 7th grades. *Participatory Educational Research, Special Issue 2016-II*, 112–120. https://doi.org/10.17275/per.16.spi.2.12

Barefoot, M. R. (2017). Student research and intrinsic motivation: Effects of formative assessment and the two-session model of information literacy instruction. *Pennsylvania Libraries: Research & Practice*, *5*(1), 13–25. https://doi.org/10.5195/palrap.2017.140

Belo, N., McKenney, S., Voogt, J., & Bradley, B. (2016). Teacher knowledge for using technology to foster early literacy: A literature review. *Computers in Human Behavior*, *60,* 372–383. https://doi.org/10.1016/j.chb.2016.02.053

Bennett, J., Gardner, R., Cartledge, G., Ramnath, R., & Council, M. (2017). Second-grade urban learners: Preliminary findings for a computer-assisted, culturally relevant, repeated reading intervention. *Education & Treatment of Children*, *40*(2), 145–186. https://doi.org/10.1353/etc.2017.0008

Bernhardt, V. L. (2016). *Data, data everywhere: Bringing all the data together for continuous school improvement* (2nd ed.). New York, NY: Routledge.

Bhagat, K. K., & Spector, M. J. (2017). Formative assessment in complex problem-solving domains: The emerging role of assessment technologies. *Journal of Educational Technology & Society*, *20*(4), 312–317.

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice*, *5*(1), 7e74. https://doi.org/10.1080/0969595980050102

Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability*, *21*(1), 5–31. https://doi.org/10.1007/s11092-008-9068-5

Boumediene, A., & Hamzaoui-Elachachi, H. (2017). The effects of formative assessment on Algerian secondary school pupils' text comprehension. *Arab World English Journal*, *8*(3), 172–190. https://doi.org/10.24093/awej/vol8no3.12

Buelin, J., Ernst, J. V., Clark, A. C., Kelly, D. P., & DeLuca, V. W. (2019). Formative evaluation techniques. *Technology and Engineering Teacher*, *78*(5), 21–23. https://www.iteea.org/39191.aspx

Bulat, J., Dubeck, M., Green, P., Harden, K., Henny, C., … Sitabkhan, Y. (2017). What works in early grade literacy instruction (knowledge and practice in international development, 1). *RTI Press Publication No. OP-0039-1702.* Research Triangle Park, NC: RTI Press. https://doi.org/10.3768/rtipress.2017.op.0039.1702

Bulunuz, N., Bulunuz, M., Karagöz, F., & Tavsanli, Ö. F. (2016). Achievement levels of middle school students in the standardized science and technology exam and formative assessment probes: A comparative study. *Journal of Education in Science, Environment and Health*, *2*(1), 33–50. https://doi.org/10.21891/jeseh.07449

Buysse, V., Peisner-Feinberg, E., Soukakou, E., Fettig, A., Schaaf, J., & Burchinal, M. (2016). Using Recognition & Response (R&R) to improve children's language and literacy skills: Findings from two studies. *Early Childhood Research Quarterly, 36*, 11–20. https://doi.org/10.1016/j.ecresq.2015.11.005

Calfee, R., Wilson, K. M., Flannery, B., & Kapinus, B. A. (2014). Formative assessment for the common core literacy standards. *Teachers College Record*, *116*(11), 1–18. https://www.tcrecord.org/content.asp?contentid=17649

Campbell, D. T., & Stanley, J. C. (1963). *Experimental and quasi-experimental designs for research*. Boston, MA: Houghton Mifflin.

Campbell, L. O., Lambie, G. W., Sutter, C. C., Bickham, A. R., & Pulse, L. P. (2018). *Measuring the predictability of Istation Indicators of Progress Early Reading (ISIP-ER) scores on Florida Standards Assessment (FSA) scores*. Orlando, FL: University of Central Florida.

Clemens, N., Hagan-Burke, S., Luo, W., Cerda, C., Blakely, A., … Jones, M. (2015). The predictive validity of a computer-adaptive assessment of kindergarten and first-grade reading skills. *School Psychology Review*, *44*(1), 76–97. https://doi.org/10.17105/spr44-1.76-97

Clinchot, M., Ngai, C., Huie, R., Talanquer, V., Lambertz, J., Banks, G., & Sevian, H. (2017). Better formative assessment. *Science Teacher*, *84*(3), 69–75. https://doi.org/10.2505/4/tst17_084_03_69

Cloonan, A., Hutchison, K., & Paatsch, L. (2016). Renewing assessment practices: Literacy teaching and learning in digital environments. *International Journal of Assessment & Evaluation*, *23*(4), 13–28. https://doi.org/10.18848/2327-7920/cgp/v23i04/13-28

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*(1), 155–159. https://doi.org/10.1037/0033-2909.112.1.155

Cotton, D. (2017). Teachers' use of formative assessment. *Delta Kappa Gamma Bulletin*, *83*(3), 39–51.

Council, M., Cartledge, G., Green, D., Barber, M., & Gardner, R. (2016). Reducing risk through a supplementary reading intervention: A case study of first- and second-grade urban students. *Behavioral Disorders, 41*(4), 241–257. https://doi.org/10.17988/bedi-41-04-241-257.1

Creswell, J. (2009). *Research design: Qualitative, quantitative, and mixed methods approaches* (Laureate Education, Inc., custom ed.). Thousand Oaks, CA: Sage.

Curry, K., Mwavita, M., Holter, A., & Harris, E. (2016). Getting assessment right at the classroom level: Using formative assessment for decision making. *Educational Assessment, Evaluation and Accountability*, *28,* 89–104. https://doi.org/10.1007/s11092 -015-9226-5

Das, S., Alsalhanie, K. M., Nauhria, S., Joshi, V. R., Khan, S., & Surender, V. (2017). Impact of formative assessment on the outcome of summative assessment: A feedback based cross sectional study conducted among basic science medical students enrolled in MD program. *Asian Journal of Medical Sciences*, *8*(4), 38–43. https://doi.org/10.3126/ajms.v8i4.17161

De Lisle (2015). The promise and reality of formative assessment practice in a continuous assessment scheme: The case of Trinidad and Tobago. *Assessment in Education: Principles, Policy & Practice*, *22*(1), 79-103. https://doi.org/10.1080/0969594X.2014.944086

Dixson, D. D., & Worrell, F. C. (2016). Formative and summative assessment in the classroom. *Theory Into Practice*, *55*(2), 153–159. https://doi.org/10.1080/00405841.2016.1148989

Dupont, P. (2018). Assessing adolescent reading comprehension in a French middle school: Performance and beliefs about knowledge. *Australian Journal of Teacher Education*, *43*(7), 30–61. https://doi.org/10.14221/ajte.2018v43n7.3

Faber, J. M., & Visscher, A. J. (2018). The effects of a digital formative assessment tool on spelling achievement: Results of a randomized experiment. *Computers & Education*, *122*, 1–8. https://doi.org/10.1016/j.compedu.2018.03.008

Faul, F., Erdfelder, E., Lang, A., & Buchner, A. (2007). *GPower* (Version 3.1.9.3) [Computer software]. http://www.gpower.hhu.de/

Fenty, N., Mulcahy, C., & Washburn, E. (2015). Effects of computer-assisted and teacher-led fluency instruction on students at risk for reading failure. *Learning Disabilities: A Contemporary Journal*, *13*(2), 141–156. http://www.ldw-ldcj.org/

Gallagher, K. (2016). *What data can do: The teacher's view of digital tools for formative assessment*. Retrieved from http://www.nasbe.org/wp-content/uploads/Gallagher.pdf

Genlott, A. A., & Grönlund, Å. (2016). Closing the gaps: Improving literacy and mathematics by ICT-enhanced collaboration. *Computers & Education*, 9968–9980. https://doi.org/10.1016/j.compedu.2016.04.004

Graham, S., Hebert, M., & Harris, K. R. (2015). Formative assessment and writing: A meta analysis. *Elementary School Journal*, *115*(4), 523–547. https://doi.org/10.1086/681947

Great Schools Partnership. (2014). *Formative assessment definition*. https://www.edglossary.org/formative-assessment

Grosas, A. B., Raju, S. R., Schuett, B. S., Chuck, J., & Millar, T. J. (2016). Determining if active learning through a formative assessment process translates to better performance in summative assessment. *Studies in Higher Education*, *41*(9), 1595–1611. https://doi.org/10.1080/03075079.2014.988704

Gustafson, S., Nordstrom, T., Andersson, U. B., Falth, L., & Ingvar, M. (2019). Effects of a formative assessment system on early reading development. *Education*, *1*, 17–27. http://www.diva-portal.org/smash/record.jsf?pid=diva2%3A1280321&dswid=-4296

Herman, J., Osmundson, E., Dai, Y., Ringstaff, C., & Timms, M. (2015). Investigating the dynamics of formative assessment: Relationships between teacher knowledge, assessment practice and learning. *Assessment in Education: Principles, Policy & Practice*, *22*(3), 344–367. https://doi.org/10.1080/0969594x.2015.1006521

Hooley, D., & Thorpe, J. (2017). The effects of formative reading assessments closely linked to classroom texts on high school reading comprehension. *Educational Technology Research & Development*, *65*(5), 1215–1238. https://doi.org/10.1007/s11423-017-9514-5

Huang, S.-C. (2016). No longer a teacher monologue: Involving EFL writing learners in teachers' assessment and feedback processes. *Taiwan Journal of TESOL*, *13*(1), 1–31. http://www.tjtesol.org/

Karim, B. (2015). The impact of teachers' beliefs and perceptions about formative assessment in the university ESL class. *International Journal of Humanities, Social Studies, and Education*, *2*, 108–115. https://ijehss.com/

Keyes, S. E., Cartledge, G., Gibson L. Jr, & Robinson-Ervin, P. (2016). Programming for generalization of oral reading fluency using computer-assisted instruction and changing fluency criteria. *Education and Treatment of Children*, *39*(2), 141–172. https://doi.org/10.1353/etc.2016.0011

Klute, M., Apthorp, H., Harlacher, J., & Reale, M. (2017). Formative assessment and elementary school student academic achievement: A review of the evidence. *Regional Educational Laboratory Program*. https://ies.ed.gov/ncee/edlabs/regions/central/pdf/REL_2017259.pdf

Laerd Statistics. (2019). *One-way repeated measures ANOVA*. Retrieved from https://statistics.laerd.com/premium/spss/ftwrma/one-way-repeated-measures-anova-in-spss.php

Lau, A. M. S. (2016). "Formative good, summative bad?" A review of the dichotomy in assessment literature. *Journal of Further & Higher Education*, *40*(4), 509–525. https://doi.org/10.1080/0309877x.2014.984600

Leech, N. L., Barrett, K. C., Morgan, G. A., & Clay, J. N., & Quick, D. (Collaborators). (2005). *SPSS for intermediate statistics: Use and interpretation* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates Publishers..

Li, H. (2016). How is formative assessment related to students' reading achievement? Findings from PISA 2009. *Assessment in Education: Principles, Policy & Practice*, *23*(4), 473–494. https://doi.org/10.1080/0969594x.2016.1139543

Luckin, R., Clark, W., Avramides, K., Hunter, J., & Oliver, M. (2016). Using teacher inquiry to support technology-enhanced formative assessment: A review of the

literature to inform a new method. *Interactive Learning Environments, 25*(1), 85–97. doi:10.1080/10494820.2015.1121152

Luo, T., Lee, G. L., & Molina, C. (2017). Incorporating Istation into early childhood classrooms to improve reading comprehension. *Journal of Information Technology Education: Research, 16,* 247–266. https://doi.org/10.28945/3788

Marin, R. (2015). *The impact of Istation reading program on reading achievement of third grade students: A mixed methods inquiry* [Doctoral dissertation, Texas A&M University-Corpus Christi]. https://tamucc-ir.tdl.org/tamuccir/bitstream/handle/1969.6/656/marin,%20rosemary%20dissertation.pdf?sequence=1

Martin, C. S., Polly, D. D., Chuang, W., Lambert, R. G., & Pugalee, D. K. (2016). Perspectives and practices of elementary teachers using an internet-based formative assessment tool: The case of assessing mathematics concepts. *International Journal for Technology in Mathematics Education, 23*(1), 3–12. https://doi.org/10.1564/tme_v23.1.01

Matthews, R., & Noyes, A. (2016). To grade or not to grade: Balancing formative and summative assessment in post-16 teacher trainee observations, *Journal of Further and Higher Education*, *40*(2), 247–261. https://doi.org/10.1080/0309877x.2014.953456

McDowell, L., Smailes, J., Sambell, K., Sambell, A., & Wakelin, D. (2008). Evaluating assessment strategies through collaborative evidence-based practice: Can one tool

fit all? *Innovations in Education & Teaching International*, *45*(2), 143–153. https://doi.org/10.1080/14703290801950310

New Jersey Department of Education. (2015). *Parent PARCC questions answered*. https://www.nj.gov/education/assessment/PARCCFAQ.pdf

No Child Left Behind Act of 2001, P.L. 107–110, 20 U.S.C. § 101. (2001). Retrieved from http://www.ed.gov/policy/elsec/leg/esea02/index.html

Owen, L. (2016). The impact of feedback as formative assessment on student performance. *International Journal of Teaching and Learning in Higher Education*, *28*(2), 168–175. http://www.isetl.org/ijtlhe/pdf/IJTLHE2158.pdf

Ozan, C., & Kincal, R. Y. (2018). The effects of formative assessment on academic achievement, attitudes toward the lesson, and self-regulation skills. *Educational Sciences: Theory and Practice*, *18*(1), 85–118. https://doi.org/10.12738/estp.2018.1.0216

PARCC Communications Team. (2014). *Did you know? What is construct validity?* Retrieved from http://www.parcconline.org/news-and-video/272-did-you-know-what-is-constructvalidity

Patarapichayatham, C. (2014). *Istation reading growth study Grades 1–8.* http://www.istation.com/Content/downloads/studies/G1-8_TX_Growth.pdf

Patarapichayatham, C. (2018). *Predictability study of ISIP Reading and Kansas Assessment Program: 3rd–6th grade students*. https://www.istation.com/Content/downloads/studies/KSPredictabilitystudy_12-2017.pdf

Patarapichayatham, C., Fahle, W., & Roden, T. R. (2014). *Predictability study of ISIP reading and STAAR reading: Prediction bands.* https://www.istation.com/Content/downloads/studies/GISD_Prediction-Band_Mar2014.pdf

Pearson. (2017). *PARCC final technical report for 2016 administration*. https://eric.ed.gov/?id=ED599197

Petour, M. (2015). Systems, ideologies and history: A three-dimensional absence in the study of assessment reform processes. *Assessment in Education: Principles, Policy & Practice*, *22*(1), 326. https://doi.org/10.1080/0969594X.2014.943153

Pinger, P., Rakoczy, K., Besser, M., & Klieme, E. (2018). Implementation of formative assessment: Effects of quality of programme delivery on students' mathematics achievement and interest. *Assessment in Education: Principles, Policy & Practice*, *25*(2), 160–182. https://doi.org/10.1080/0969594x.2016.1170665

Polly, D., Wang, C., Martin, C., Lambert, R. G., Pugalee, D. K., & Middleton, C. W. (2017). The influence of an internet-based formative assessment tool on primary grades students' number sense achievement. *School Science & Mathematics*, *117*(3/4), 127–136. https://doi.org/10.1111/ssm.12214

Putman, R. S. (2016). Technology versus teachers in the early literacy classroom: An investigation of the effectiveness of the Istation integrated learning

system. *Educational Technology Research and Development, 65*(5), 1153–1174. https://doi.org/10.1080/0969594x.2016.1170665

Rakoczy, K., Pinger, P., Klieme, E., Hochweber, J., Schütze, B., & Besser, M. (2019). Formative assessment in mathematics: Mediated by feedback's perceived usefulness and students' self-efficacy. *Learning & Instruction, 60,* 154–165. https://doi.org/10.1016/j.learninstruc.2018.01.004

Ramaprasad, A. (1983). On the definition of feedback. *Behavioral Science, 28,* 4–13. https://www.mdpi.com/journal/behavsci

Ravenel, J., Lambeth, D. T., & Spires, B. (2014). Effects of computer-based programs on mathematical achievement scores for fourth-grade students. *Journal on School Educational Technology*, *10*(1), 8–21. https://doi.org/10.26634/jsch.10.1.2830

Sach, E. (2015). An exploration of teachers' narratives: What are the facilitators and constraints which promote or inhibit "good" formative assessment practices in schools? *Education, 43*(3), 322–335. https://doi.org/10.1080/03004279.2013.813956

Sadler, R. (1989). Formative assessment and the design of instructional systems. *Instructional Science, 18,* 119–144. https://doi.org/10.1007/bf00117714

Sadler, R. (1998). Formative assessment: Revisiting the territory. *Assessment in Education: Principles, Policy & Practice*, *1*(5), 77–84. https://doi.org/10.1080/0969595980050104

Saito, H., & Inoi, S. (2017). Junior and senior high school EFL teachers' use of formative assessment: A mixed-methods study. *Language Assessment Quarterly, 14*(3), 213–233. https://doi.org/10.1080/15434303.2017.1351975

Shadish, W., Cook, T., & Campbell, D. (2002). *Experimental and quasi-experimental designs for generalized causal inference.* Boston, MA: Wadsworth Cengage Learning.

Shirley, M. L., & Irving, K. E. (2015). Connected classroom technology facilitates multiple components of formative assessment practice. *Journal of Science Education and Technology*, *24*(1), 56–68. https://doi.org/10.1007/s10956-014-9520-x

Shute, V., & Rahimi, S. (2017). Review of computer-based assessment for learning in elementary and secondary education. *Journal of Computer Assisted Learning, 33*(1), 1–19. https://doi.org/10.1111/jcal.12172

Simmons, D. C., Kim, M., Kwok, O., Coyne, M. D., Simmons, L. E., Oslund, E., & Rawlinson, D. (2015). Examining the effects of linking student performance and progression in a Tier 2 kindergarten reading intervention. *Journal of Learning Disabilities*, *48*(3), 255–270. https://doi.org/10.1177/0022219413497097

Spector, J., Ifenthaler, D., Sampson, D., Yang, L., Mukama, E., Warusavitarana, A., & Gibson, D. (2016). Technology enhanced formative assessment for 21st century learning. *Journal of Educational Technology & Society*, *19*(3), 58–71. https://www.j-ets.net/collection/published-issues/19_3

Srivastava, T. K., Waghmare, L. S., & Mishra, V. (2018). Formative assessment classroom techniques (FACTs) for better learning in pre-clinical medical education: A controlled trial. *Journal of Clinical & Diagnostic Research*, *12*(9), 1–8. https://doi.org/10.7860/jcdr/2018/35622.11969

Tabachnik, B. G. & Fidell, L. S. (2007). *Using multivariate statistics* (5th ed.). Needham Heights, MA: Allyn & Bacon.

Texas Success Training. (2013). *Istation reading.* Retrieved from http://www.bisd.net/cms/lib02/TX01001322/Centricity/Domain/41/Texas%20SUCCESS %20Software%20Training.pdf

Timmis, S., Broadfoot, P., Sutherland, R., & Oldfield, A. (2015). Rethinking assessment in a digital age: Opportunities, challenges and risks. *British Educational Research Journal, 42*(3), 454–476. https://doi.org/10.1002/berj.3215

Trochim, W. M., Donnelly, J. P., & Arora, K. (2016). *Research methods: The essential knowledge base.* https://cengagebrain.vitalsource.com/#/books/9781305445185/cfi/51!/4/4

U.S. Department of Education, Office of Elementary and Secondary Education, Office of State Support. (2015). *Improving basic programs operated by local educational agencies* (Title I, Part A).

Weinfurt, K. P. (2000). Repeated measures analyses: ANOVA, MANOVA, and HLM. In L. G. Grimm & P. R. Yarnold (Eds.), *Reading and understanding more multivariate statistics* (pp. 317–361). Washington, DC: American Psychological Association.

Xu, Y., & Brown, G. (2016). Teacher assessment literacy in practice. *Teaching and Teacher Education*, *58*, 149–162. https://doi.org/10.1016/j.tate.2016.05.010

Zlatovic, M., Balaban, I., & Kermek, D. (2015). Using online assessments to stimulate learning strategies and achievement of learning goals. *Computers & Education, 91*, 32–45. http://www.elsevier.com/locate/compedu