The Relationship Between iReady Diagnostic and Mock Interim Assessments on Grades 3-5 Student Performance on TCAP Scores

By

Amy A. McAmis

A dissertation submitted in partial fulfillment of the requirements for the degree of

Doctor of Education

Education Department

School of Sciences and Education

Milligan University, Tennessee

2022

Doctoral Committee:

Major Professor

Dr. Patrick Kariuki

Dr. Mark Dula

Dr. Janet Faulk

Abstract

The purpose of this quantitative study was to examine the relationship between iReady Spring Diagnostic for Tennessee and Tennessee Mock Interim Assessment scores on students Grades 3-5 Tennessee Comprehensive Assessment Program (TCAP) ELA and math scores. Participants included 995 students from seven elementary schools in an Upper East Tennessee school district who were administered the iReady Spring Diagnostic Assessments, the Tennessee Mock Interim Assessments, and TCA P TNReady tests for ELA and math in Grades 3-5 in the 2020-2021 school year. Six major research questions and their subset questions and the null hypotheses were analyzed using Pearson correlation coefficient and multiple linear regression. The results indicated a significant relationship between iReady Spring Diagnostic and TCAP TNReaady test scores in ELA and math for Grades 3-5. Similarly, there was a significant relationship between Mock Interim Assessment scores and TCAP TNReady test scores in ELA and math for Grades 3-5. Multiple linear regression test revealed that iReady Spring Diagnostic Assessments and Mock Interim Assessments were both significant predictors of TCAP TNReady test scores in ELA and math for Grades 3-5. However, iReady Spring Diagnostic Assessments were stronger predictors of ELA Grades 3-5 and math for Grades 3 and 5 than Mock Interim Assessments, while Mock Interim Assessments were stronger predictors of TCAP TNReady math test scores for Grade 4 than iReady Spring Diagnostic Assessments. The results suggest that both iReady Spring Diagnostic Assessments and Tennessee Mock Interim Assessments are useful formative assessment tools for ELA and math.

Keywords: Formative Assessment, Interim Assessment, Computerized Adaptive Test, Tennessee Comprehensive Assessment Program (TCAP), iReady Diagnostic Assessment, Mock Interim Assessment

Milligan University Dissertation Defense Approval Form

Candidate Name: Amy McAmis

Date of Defense: 4/19/22

Dissertation Title: The Relationship Between iReady Diagnostic and Mock Interim Assessments on Grades 3-5 Student Performance on TCAP Scores

Final Approval of Dissertation Defense

A signature below indicates committee members agree with the following:

1. Agreement the dissertation meets with the committee's approval.
2. Agreement the oral defense of the dissertation was successful.

Patricia M. Kanuckel — 4/19/22

Dissertation Chair Signature — Date

[signature] — 4/19/22

Dissertation Committee Member Signature — Date

[signature] — 4/19/22

Dissertation Committee Member Signature — Date

Other Required Signatures

Patricia M. Kanuckel — 4/19/22

Ed. D. Research Director Signature — Date

Angela Dalton-Puffenbarger — 04/19/2022

Area Chair of Education signature — Date

Dedication

Blessed is the man that trusteth in the LORD, and whose hope the LORD is.

Jeremiah 17:7

This research study is dedicated…

To God through Jesus Christ from whom all things are possible to those who are steadfast in Him through faith, hope, and love in the valleys and on the mountaintops for your ways are higher than mine.

To my husband, Scott. You have supported, encouraged, and believed in me when I doubted myself. I am so very thankful and blessed that you have shared this life-long journey of learning with me. Thank you for saying, "just do it" and pushing me to fulfill a goal and dream I set many years ago. I think it's time for us to take a vacation with no computers allowed! Love you!

To my children, Megan and Micah. I know God has great plans for both of you. My prayer is that you always seek God first and allow Him to take care of the rest! Never quit dreaming nor give up on your dreams for with God all things are possible. I can't express how thankful I am for your love, patience, and support each time I have gone back to school. You will always be my babies and I love you both with all my heart!

To my grandchildren. I know God has great plans for each of you. My prayer is that each of you knows how much you are loved by God and your family! And, to dream big dreams for with God all things are possible! Guess what? Nana is done with school and it's time to have some fun! I love you all!

To my mother, Carolyn Booker. You have been my biggest cheerleader. Your love support, and prayers are priceless! You constantly checked on me and made sure that I was taking care of myself. Thank you for all the sacrifices you have made along the way that has led to making this joyous time possible for me. Thank you for showing me and teaching me that a mother's love is endless! I love you, mom!

To my extended family, Scott and Amy Booker, Wesley Booker, and Eileen McAmis. I am so grateful to be a part of such a wonderful and loving family. Thank each of you for your prayers, love, and support. I love you all!

In memory of my father, Dennis Booker. I love you and miss you dearly! I know you would be here with the biggest smile and sharing this time with me if you could. However, your silent strength, unwavering love, strong work ethic, and steadfast faith in God remain strongly in my heart and always will. I miss you, daddy!

In memory of my grandmother, Mable Booker. You have been such an inspiration to me all my life. I would not be where I am today without your love of learning, support, and countless trips to ETSU during my undergraduate program. I wish you were here to share this time with me. Yet, that would be selfish of me when you were the least selfish person I have ever known. I miss you and will always love you!

I am truly grateful and blessed for the love and support from my family and thankful that God directs my steps!

A man's heart plans his way, But the Lord directs his steps.

Proverbs 16:9

Acknowledgments

I would like to extend my deepest gratitude to all my family and friends who have supported and encouraged me throughout my educational journey and this doctorate program. I would also like to extend my sincere gratitude to those special teachers along my educational journey who have shared their love of learning and inspired me to follow God's path.

I would like to extend my sincere gratitude to Dr. Patrick Kariuki for serving as my dissertation chair. I would like to thank Dr. Patrick Kariuki for sharing your educational expertise, wisdom, and faith. Thank you for your guidance, patience, and support. Thank you for being selfless with your time during my dissertation process. I appreciate and admire your commitment to the success of Milligan students. Thank you for being my lighthouse!

I would like to extend my sincere gratitude to Dr. Janet Faulk for serving on my dissertation committee. I would like to thank Dr. Janet Faulk for editing my drafts with kindness and supportive feedback. I appreciate your recommendations and keen eye that bettered my writing skills. Thank you!

I would like to extend my sincere gratitude to Dr. Mark Dula for serving on my dissertation committee. I would like to say a special thank you to Dr. Mark Dula for your genuine willingness to help me whenever called upon. I appreciate and am in awe of your mathematical mind and knowledge. Thank you for showing me that I could do quantitative statistics and it could be fun, relatively speaking. Your continuous support was invaluable throughout my entire doctorate program. Thank you!

I would like to extend further gratitude to Dr. David Timbs, Dr. Angela Hilton-Prillhart, Dr. Jacki Wolfe, and Dr. Tausha Clay for sharing your enthusiasm of and for learning and Milligan University, respectively. It was contagious. I would like to say a special thank you to

Dr. David Timbs. Thank you, Dr. David Timbs, for sharing your love of learning with me as well as for pushing, pulling, and stretching me to new heights. You truly represent a servant leader and a life-long learner and the joy that comes with them. Thank each of you!

I would like to express my thankfulness to the studied school district. Thank you for supporting me and this work when called upon with kindness and expediency. I hope this work is beneficial in helping present and future students and educators. Helping others to attain quality education for a brighter future is what it's all about! Thank you!

I would like to extend my deepest gratitude to my colleagues and co-workers for your support. I would like to say a special thank you to Lori Loftis for being the special person, teacher, colleague, and friend you are to me and those around you. Thank you for your support and help as I embarked on this personal and professional journey!

I would also like to express my thankfulness to all my past, present, and future students for inspiring me to be an educator! Thank you!

Last but not least, I would like to say an extra special thank you to my cohort! I entered this program knowing no one from Milligan or the surrounding area. I am leaving grateful with five new very special friends in my life! God put together a small but powerful group-Cohort 4! Thank you, Dr. Sarah Gillette, Dr. Jeff Gray, Dr. Christina Maiden, Dr. Daisy Sanders, and Leeanne Francis for your never-ending support, wit, humor, encouragement, and friendship…we have laughed, cried, vented, celebrated, and survived a pandemic and a doctorate program together! I am truly blessed to have gotten to know each of you and now call you friends! Leeanne your turn! We made it through in 2022!

Table of Contents

List of Tables

CHAPTER 1

Introduction

In the journey of preparing students to be college and career-ready to compete in the 21st-century global economy, every educational system has the potential to reach greater success within their educational landscape. According to Hall and Simeral (2008), this statement can be made "because every school is full of children, who possess limitless potential" (p. 3). The potential for greater success includes increasing student success rates, closing achievement gaps, meeting the individual needs of students, and creating successful, confident, competent, and prepared students (Hall & Simeral, 2008). The quality of teaching and learning and leadership in schools are the main influencers in improving student achievement for all students; therefore, teachers and leaders are corded together for the advancement of student learning and achievement outcomes. In a time of strenuous accountability policies focusing on student learning and student achievement, educators are continuously being scrutinized for instructional decisions for procuring student achievement. The "use of data has become more central to how many educators evaluate their practices and monitor students' academic progress" (The Institute of Education Sciences, 2009, p. 1). Data and its availability are some of the most powerful resources available to educators to enhance instructional decisions intended to increase student learning and achievement.

Accountability policies from federal, state, and local stakeholders to improve student achievement have thrust educational systems into an age of accountability and into an era driven by political goals aimed at educational reform where excellence in education is often equated to high-stakes test scores. Increasing student achievement has been at the forefront of educational reform since the initial passage of *the Elementary and Secondary Act* of 1965 (*ESEA*), which has

been accelerated by the reauthorization of the *Elementary and Secondary Act* of 1965 called the *No Child Left Behind Act* of 2001 (NCLB) and *Every Student Succeeds Act* of 2015 (United States Department of Education, 2004; United States Department of Education, 2015). Under NCLB, states were forced to set academic standards that moved from norm-based testing to criterion-based testing, which focused on achievement levels over time through high-stakes testing and gave the government unprecedented authority when schools failed to meet NCLB's performance targets (United States Department of Education, 2004). Under the Obama administration, *A Blueprint for Educational Reform* built upon the core principles of NCLB and held states to rigorous standards by developing and implementing college and career-ready standards with aligned assessments and required states to provide intensive support and interventions (United States Department of Education, 2010).

The *Race to the Top* (RTT) initiative sought to continue the Obama administration's reform efforts to ensure all students were college and career ready through a grant competition to selected states, which "agreed to implement a range of education policies and practices designed to improve student outcomes" (U.S. Department of Education, 2016, p. 5). Tennessee received over 500 million dollars of RTT funds, which created an unprecedented and intensive focus on adopting new standards, assessments, and accountability and data systems to measure student growth and success (Tennessee State Government, 2010b). It also included a teacher development component with emphasis on strategies for improving instruction.

This set the tone for school systems to educate all students to high levels of achievement, which is measured through high-stakes summative testing. School systems are continuously seeking avenues such as formative assessment systems to provide data to increase student learning and student achievement throughout the teaching and learning cycle as measured by test

scores through high stakes annual testing such as the Tennessee Comprehensive Assessment Program (TCAP). TCAP is the Tennessee Department of Education's requirement for end-of-the-year summative assessments for grades three through eight (Tennessee Department of Education, n.d.e). The purpose of TCAP assessments is to assess students' skill level and understanding of the Tennessee Academic Standards and provide a summative measure of achievement in reading, math, science, and social studies as well as provide reports on student performance levels to school districts, teachers, and parents (Tennessee Department of Education, n.d.e). These assessments are administered in compliance with the *Every Student Succeeds Act of* 2015 and T.C.A. § 49-1-602 for district and school accountability (Tennessee Department of Education, n.d.e).

Educators can read about their schools and school districts' successes or failures in the newspaper or online. Public reporting of yearly mandated state standardized test scores reflects schools' ability to increase student learning and achievement. This necessitates a priority of focus on program evaluation of computerized adaptive assessments used to monitor student learning, achievement, and predictability of student performance level on high-stakes tests. According to Fitzpatrick et al. (2011), mainstream program evaluation is already within many schools today and is considered "part of the organization's work ethic, its culture, and job responsibilities at all levels of the organization" (p. 236). As defined by Fitzpatrick et al. (2011), mainstream program evaluation "is the process of making evaluation an integral part of an organization's everyday operations" (p. 236). Evaluation provides empowerment in internal decision-making and taking action on program planning and improvement as related to the teaching and learning cycle (Fitzpatrick et al., 2011) and is often equated to high-stakes summative state test scores. Educational leaders at all levels need to know if their formative computerized adaptive

assessment systems are producing the data and results claimed by vendors to make instructional decisions for optimal student learning and achievement that is measured on high-stakes state summative tests. Educators at every level are highly invested in student scores as they relate to many educational aspects ranging from educator level of effectiveness to school funding. Educators want their students to grow, learn, and experience success on the assessments and be adequately prepared for their future career path of choice.

Background of the Problem

Although assessments have been a part of the educational system for more than 100 years, educational assessment methods and formats have changed over time (Shepard, 2016). The purpose of using assessments to measure student achievement, evaluate student progress, and inform instruction is a common, well-established practice in education and remains at the forefront in educators' decision-making. It is vital to assess students' performance and progress throughout the school year and utilize the data to tailor instruction. With national and state legislation pressuring schools and districts to raise state summative test scores, this places an enormous challenge on educators to find and utilize formative assessments that will provide adequate and accurate data of student performance and progress.

School districts often opt to use a variety of online computerized assessment systems such as iReady Diagnostic and Mock Interim Assessments for data-based evidence to measure instructional effectiveness, student learning, and to improve student performance levels on state exams such as the TCAP. According to a study by Curriculum Associates (2020) done in partnership with the Educational Research Institute of America (ERIA), there is a strong correlation between iReady Diagnostic Assessment for Tennessee test scores and TCAP test scores. iReady assessment system provides a suite of data reports that transform data into

meaningful and actionable insights for differentiating instruction through pinpointing student strengths and knowledge gaps with targeted instructional tools (Curriculum Associates, 2021a). Tennessee Department of Education (2020b) recently released the Mock Interim Assessments that "mirrors the current TCAP summative assessments, as well as provide accurate scaled scores and performance bands that will estimate each student's performance" (p.12). The Mock Interim Assessments is part of the Schoolnet Instructional Improvement System, which provides a dashboard for educators at various levels in a school system to make data-informed decisions regarding relevant data analysis, tasks, and resources (Pearson, 2015). When considering research-based programs, materials, or products as a consumer, educational leaders must seek the effectiveness and purpose of the product, while looking past the research-based advertisement and examining the research behind the research, the validity, and reliability of the program, material, or product.

Statement of the Problem

Although accountability trends may explain why more available data is in schools, the question often remains of which type of assessment produces quality data to assess what students are learning and to what degree students are progressing toward goals and higher performance levels on annual high stakes summative tests such as the TCAP. The need for educational systems and educators to identify the right type of online formative assessment system can be daunting as pressure continues to mount through high-stakes testing within the realm of accountability and political ramifications. Most people think in terms of the outcomes in regards to product or program evaluation, especially when advertised as researched-based, research-tested, or research-proven; therefore, determining the criteria for product or program evaluation is reliant on the product’s intent and the purpose of the evaluation in determining the scope,

subject, and methods (Fitzpatrick et al., 2011). According to the National Research Council of Teachers of English (2014), an effective assessment system should employ multiple assessments, be representative of standardized tests, and ensure tests are valid and reliable for student populations. The product or program evaluation process should be rigorous, systematic, and ethical (Fitzpatrick et al., 2011).

While there is evidence that the use of formative assessments guide and inform instruction that leads to gains in student learning and student achievement (Black & Wiliam, 1998a; 1998b), it is necessary to examine the quality of formative computerized assessments and testing data for assessment systems such as iReady Diagnostic Assessments for Tennessee and the Tennessee Mock Interim Assessments. School leaders must have accurate and up-to-date information to make better-informed decisions in their quest for online formative assessment systems aligned to state academic standards that will improve student performance levels to meet state academic standards assessed on annual summative state exams. Therefore, this research will examine what the data suggest to assist educational leaders' decision-making process to select the better online formative assessment tools based on student performance levels between iReady Diagnostic Assessments for Tennessee and the Tennessee Mock Interim Assessments in regards to student performance on TCAP test scores.

Purpose of the Study

The purpose of this quantitative research study was to examine the relationship between the scores of 3rd – 5th-grade students on iReady Spring Diagnostic for Tennessee (iReady Diagnostic) and the Tennessee Mock Interim Assessments (Mock Interim Assessment) and TCAP TNReady summative test scores for ELA and math. A related purpose of this study was

to determine the predictive validity of iReady Diagnostic and Mock Interim Assessments scores for ELA and math on Grades 3-5 TCAP TNReady ELA and math performance.

Research Questions and Null Hypotheses

Six major research questions and their subset questions including the null hypotheses were used to guide the collection and analysis of data for this research.

1. Is there a significant relationship between iReady Diagnostic scores and 3rd-grade TNReady ELA test scores?

   $H_0$1: There is no significant relationship between iReady Diagnostic scores and 3rd-grade TNReady ELA test scores.

1a. Is there a significant relationship between Mock Interim scores and 3rd-grade TNReady ELA test scores?

   $H_0$1a: There is no significant relationship between Mock Interim scores and 3rd-grade TNReady ELA test scores.

1b: Is iReady Diagnostic Assessments or Mock Interim Assessments the best predictor of 3rd-grade TCAP ELA test scores?

   $H_0$1b: There is no best predictor of 3rd-grade TCAP ELA test scores by iReady Diagnostic Assessments or Mock Interim Assessments.

2. Is there a significant relationship between iReady Diagnostic scores and 3rd-grade TNReady math test scores?

   $H_0$2: There is no significant relationship between iReady Diagnostic scores and 3rd-grade TNReady math test scores.

2a: Is there a significant relationship between Mock Interim scores and 3rd-grade TNReady math test scores?

$H_0$2a: Is there a significant relationship between Mock Interim scores and 3rd-grade TNReady math test scores?

2b: Is iReady Diagnostic Assessments or Mock Interim Assessments the best predictor of 3rd-grade TCAP math test scores?

$H_0$2b: There is no best predictor of 3rd-grade TCAP math test scores by iReady Diagnostic Assessments or Mock Interim Assessments.

3. Is there a significant relationship between iReady Diagnostic scores and 4th-grade TNReady ELA test scores?

$H_0$3: There is no significant relationship between iReady Diagnostic scores and 4th-grade TNReady ELA test scores.

3a. Is there a significant relationship between Mock Interim scores and 4th-grade TNReady ELA test scores?

$H_0$3a: There is no significant relationship between Mock Interim scores and 4th-grade TNReady ELA test scores.

3b. Is iReady Diagnostic Assessments or Mock Interim Assessments the best predictor of 4th-grade TCAP ELA test scores?

$H_0$3b: There is no best predictor of 4th-grade TCAP ELA test scores by iReady Diagnostic Assessments or Mock Interim Assessments.

4. Is there a significant relationship between iReady Diagnostic scores and 4th-grade TNReady math test scores?

$H_0$4: There is no significant relationship between iReady Diagnostic scores and 4th-grade TNReady math test scores.

4a. Is there a significant relationship between Mock Interim scores and 4th-grade TNReady math test scores?

$H_0$4a: There is no significant relationship between Mock Interim scores and 4th-grade TNReady math test scores.

4b. Is iReady Diagnostic Assessments or Mock Interim Assessments the best predictor of 4th-grade TCAP math test scores?

$H_0$4b: There is no best predictor of 4th-grade TCAP math test scores by iReady Diagnostic Assessments or Mock Interim Assessments.

5. Is there a significant relationship between iReady Diagnostic scores and 5th-grade TNReady ELA test scores?

$H_0$5: There is no significant relationship between iReady Diagnostic scores and 5th-grade TNReady ELA test scores.

5a. Is there a significant relationship between Mock Interim scores and 5th-grade TNReady ELA test scores?

$H_0$5ba: There is no significant relationship between Mock Interim scores and 5th-grade TNReady ELA test scores.

5b. Is iReady Diagnostic Assessments or Mock Interim Assessments the best predictor of 5th-grade TCAP ELA test scores?

$H_0$5b: There is no best predictor of 5th-grade TCAP ELA test scores by iReady Diagnostic Assessments or Mock Interim Assessments.

6. Is there a significant relationship between iReady Diagnostic scores and 5th-grade TNReady math test scores?

$H_0$6: There is no significant relationship between iReady Diagnostic scores and 5th-grade TNReady math test scores.

6a. Is there a significant relationship between Mock Interim scores and 5th-grade TNReady math test scores?

$H_0$6a: Is there a significant relationship between Mock Interim scores and 5th-grade TNReady math test scores?

6b. Is iReady Diagnostic Assessments or Mock Interim Assessments the best predictor of 5th-grade TCAP math test scores?

$H_0$6b: There is no best predictor of 5th-grade TCAP math test scores by iReady Diagnostic Assessments or Mock Interim Assessments.

Significance of the Study

Many school districts use online formative computerized assessment systems such as iReady Diagnostic for Tennessee and the Tennessee Mock Interim Assessments to evaluate the teaching and learning cycle, student learning, achievement, and to inform and drive instruction. Therefore, it is imperative to gather and analyze data surrounding the use of these online formative assessments on student performance levels on the TCAP tests. District leaders, school leaders, and teachers continuously are making crucial decisions through data gleaned from these assessments. Educators must understand the rationale behind formative assessment research (Popham, 2006b). Given the pressure on educators to raise student performance levels on high-stakes accountability tests and the attraction of using formative assessments by educational institutions, some test publishers began relabeling many of their testing packages as ‘formative”. For assessment systems, including online formative assessments, to be truly effective at the district level, they must be able to provide different kinds of information to various policy and

decision-makers at different times in various forms for three different levels the classroom, the school, and the district (Stiggins & DeFour, 2009).

This research will add to the existing body of research on formative assessment systems, including those deemed computerized adaptive assessments, and serve as a vehicle to improve student learning, achievement, and performance level on state exams. Educators use formative computerized assessments to measure the effectiveness of the teaching and learning cycle to increase student learning and achievement more intensely due to modern accountability systems, which are based on state standardized test scores. When viewed as an approach, formative computerized assessments can be used to provide educators with real-time data-based evidence of student learning, progress, and achievement in mastering skills and knowledge assessed on state standardized tests to optimize student learning beforehand (Stiggins, 2005; Wiliam, 20014; Van der Kleij et al., 2015). Additionally, this research shows the importance for educational leaders to consider the design of an online formative assessment system that will adequately fulfill the more important purpose first. Then they can consider if the same assessment can fulfill additional purposes or if multiple assessments are needed for a comprehensive assessment system (Perie et al., 2009; Perie et al. 2007).

Leaders must continue to seek for and use online formative assessment tools and systems aligned to state academic standards to measure student learning and achievement that will improve student performance levels on academic standards assessed on state standardized tests. In that process, they should identify the intended purposes of the assessments and ensure that the assessments they select are validated to meet those purposes. While there may be existing research on formative assessment and use, it is imperative educational decision-makers have the most recent available data from online formative assessment systems currently used to make the

best decision for their respective student population regarding functionality, validity, and reliability of online formative assessment systems available.

Definition of Terms

The following definitions of terms were used in this study to ensure uniformity and understanding of the terms throughout the study. The researcher developed and defined terms not accompanied by a citation.

Computerized Adaptive Test (CAT): Computerized adaptive tests are assessments that adjust to the level of knowledge of the student based on the item response theory (IRT). IRT is based on a statistical method used to predict a student's ability to achieve a score based on unobservable latent traits such as aptitude, achievement, skill, and personality and their manifestation such as outcomes, responses, or performance (Zanon et al., 2016). IRT aims to establish a relationship link between tested items, individual responses to those tested items, and the underlying latent traits being measured (Lang & Tay, 2020). Computerized adaptive tests and computerized adaptive assessments may be used interchangeably in this study.

Formative Assessments: The Council of Chief State School Officers (CCSSO) defined formative assessment by the Formative Assessment for Student and Teachers (FAST) as a process used during instruction to provide feedback to adjust or make modifications for ongoing teaching and learning to improve student achievement as related to instructional objectives (Popham, 2006a). Popham (2008, as cited Dunn & Mulvenon, 2009) expanded his definition of formative assessment as a planned process that the teacher and student use the assessment data as evidence to modify or adjust ongoing instruction and learning.

High-stakes Testing: "In general, "high stakes" means that test scores are used to determine punishments, accolades, advancement, or compensation" (Great Schools Partnership,

2014, para. 1). Summative tests, summative assessments, state summative tests, state standardized tests, and summative end-of-the-year tests referring to high-stakes tests may be used interchangeably in this study.

Interim Assessments: Benchmark, diagnostic, formative, and or predictive assessments to provide data on improving student performance levels within the school year fall under the umbrella term 'interim assessment' (Perie et al., 2009; Perie et al, 2007).

iReady Diagnostic Assessments for Tennessee: iReady Diagnostic assessments for Tennessee are computer-adaptive assessments designed to provide educators with actionable insight into student needs and strengths and are based on Tennessee state academic standards (Curriculum Associates, 2021a). iReady assessment system provides a suite of data reports that transform data into meaningful and actionable insights for differentiating instruction through pinpointing student strengths and knowledge gaps with targeted instructional tools.

Predictive Purposes: The ability or degree of a tool, such as an assessment, to predict the future performance on an assessment of the same construct (Immekus & Atitya, 2016).

Tennessee Comprehensive Assessment Program (TCAP): TCAP is the Tennessee Department of Education's requirement for end-of-the-year summative assessments for Grades 3-8 with 2nd grade being optional (Tennessee Department of Education, n.d.e). The purpose of TCAP assessments is to assess students' skill level and understanding of the Tennessee Academic Standards and provide a summative measure of achievement in reading, math, science, and social studies as well as provide reports on student performance levels to school districts, teachers, and parents (Tennessee Department of Education, n.d.e).

Tennessee Mock Interim Assessments: The Tennessee Mock Interim Assessments are assessments that "mirror the current TCAP summative assessments, as well as provide accurate

scaled scores and performance bands that will estimate each student's performance" (p.12), which have recently been released by the Tennessee Department of Education (2020). The Mock Interim Assessments is part of the Schoolnet Instructional Improvement System, which provides a dashboard for educators at various levels in a school system to make data-informed decisions regarding relevant data analysis, tasks, and resources (Pearson, 2015).

Limitations of the Study

This study is limited because the results may only be generalized to the population within one East Tennessee school district.

Organization of the Study

Chapter One provides a general introduction to the study followed by the statement of the problem, the purpose of the study, eight research questions and null hypotheses, the significance of the study, definitions of terms, and limitations of the study. Chapter Two contains a review of literature and research that is relative to the topics and issues addressed in this study pertinent to the historical significance, impact, and use of educational assessments and the validity and reliability. The research methodology and procedures used to gather data are presented in Chapter Three. The results, analyses, and findings from the study are presented in Chapter Four. Chapter Five contains a summary of the study and findings, a discussion, conclusions from the findings, and recommendations for further study.

CHAPTER 2

Literature Review

Accountability policies from federal, state, and local stakeholders to improve student achievement have thrust educational systems into an age of accountability and into an era driven by political goals aimed at educational reform where excellence in education is often equated to high-stakes student test scores. School leaders must have accurate and up-to-date information to make better-informed decisions in their quest for effective online formative assessment systems aligned to state academic standards that will improve student performance levels to meet state academic standards assessed on annual summative state exams. This research aims to assist educational leaders' decision-making process to select the better formative computerized assessment tools and systems that will provide an adequate and accurate stream of data to measure student learning and progression toward meeting state academic standards.

Since there is minimal research available on the predictive use of student scores from formative computerized adaptive assessments to guide decision-makers at the classroom, school, and district levels (Immekus & Atitya, 2016; Sutter et al., 2020), this literature review focuses primarily on the historical and current significance, impact, and use of educational assessments for accountability purposes and measures placed on educators in public education, specifically including Tennessee. Therefore, chapter two is organized into the following sections: The Role of Government in Education and Decisions, History of Formative Assessments, Assessments for Accountability Measures in Tennessee (Standardized Tests), Formative Computerized Adaptive Tests, Local Formative Computerized Tests, Program Evaluation, and Theoretical Framework.

The Role of Government in Education and Decisions

The federal government has supported educational endeavors throughout the history of American education in a variety of ways. The federal government's role in educational initiatives, particularly after 1965, emphasizes the importance of investing in education, promoting economic growth, and increasing opportunity for children in America as well as promoting civil rights by holding the states accountable for children of all backgrounds and students with disabilities access to free and appropriate education (McShane, 2021). The current high-stakes testing trend in the United States was initiated by the enactment of the *Elementary and Secondary Education Act of 1965* under President Lyndon Johnson, which was created to make education opportunities and services more equitable for low-income students and included new accountability and testing provisions focused on raising academic standards (McGuinn, 2015). As a result of these new requirements, high-stakes testing and accountability systems were set in motion to improve education and student learning and achievement. This increased the government's role in educational decisions through policy mandates in the name of educational reform.

Educational leaders must make the best financial decisions utilizing federal, state, and local funds for the procurement of students' learning and achievement in regard to programs and products to further promote and support student learning, improvement, and achievement via current accountability systems, which is based on high-stakes student test scores tests. It is important to understand the evolution of federal and state governments' involvement and role in education and decisions pertinent to public education. History shows how the federal government influences state and local education policy creating dilemmas to meet mandated policies and forging influences in educational leaders' decision-making processes.

Pre-ESEA 1965

Public education was constitutionally in the hands of the states, not the federal government through the reserve clause in the Tenth Amendment. Therefore, the power to create school systems was left in the hands of the states. Ornstein (1984) states:

> Although the Constitution makes no mention of public education, the Tenth Amendment reserved to the states the powers not delegated to the federal government or prohibited to the states. Federal recognition of the states' responsibility for education was embodied in the Northwest Ordinances of 1785 and 1787. (p. 1)

The *Federal Land Ordinance of 1785* required land to be reserved for school buildings.

However, in 1790 state public schooling began when Pennsylvania required free education only for those who could not afford it, whereas in 1820 Massachusetts was the first state to have a tuition-free high school for all (Hornbeck, 2017). According to Webb (2005), less than 25,000 students were enrolled in public schools in 1875, yet this era marked the beginning of free, tax-supported schools. The Common School Movement aimed at providing all children with free public education had spread to most states by the late 1800s, and by 1930 every state had a compulsory education law, which led to states and local cities increasing control over schools (Hornbeck, 2017).

The first southern state to enact a compulsory school attendance law was Tennessee (Webb, 2005). All children in Tennessee between the ages of eight and sixteen were required to attend school by law in 1922. Tennessee's first sales tax law, which allotted 80% of the funds to education, passed Tennessee State Legislature in 1947 and students began receiving free textbooks in 1953. This constitutes the beginning of the Tennessee state government's role in public education, which increases as the federal government's role increases in public education.

The federal government's focus was on the disadvantaged, minorities, women, and the handicapped, and with federal funds came monitoring, reporting, auditing, and compliance, which grew after the Great Depression and World War II (Hornbeck, 2017) and substantially after 1965. Initially, federal funds came as no-strings-attached, then changed to multiple-strings-attached as the federal government's role to improve education evolved.

Post-ESEA 1965

Before the passage of the *National Defense Education Act of 1958* (NDEA) and the *Elementary and Secondary Education Act of 1965* (ESEA), education in the United States had been controlled primarily by state and local entities and the federal government had no constitutional rights or authority to demand educational policy to states (McGuinn, 2015). However, since the passage of the NDEA and ESEA the federal government's role in education accountability systems was set in motion, which set the tone for future education reform efforts by leveraging federal funds to improve public education and student learning and achievement through the measurement of high-stakes test scores.

The beginning contributions of the U. S. Congress with mandated policies using grant-in-aid funds include the *National Defense Education Act*, which gave money to localities to promote innovation; the *Civil Rights Act of 1964*, which gave educational rights to all students; the *Elementary and Secondary Education Act of 1965*, which gave money to the economically disadvantaged under Title 1; and, the *Education for All Handicapped Children Act of 1975*, which provided students with disabilities a free and appropriate public education (Superfine, 2011). Through the 1970s, the federal government's role in education was fairly limited in many aspects.

However, the 1980s saw two contradictory trends of government power at play (McGuinn, 2015). The creation of the Department of Education as a Cabinet-level agency gave the federal government more federal authority over public schools. On the other hand, the Reagan administration sought to limit federal bureaucracy thus limiting federal authority to hold states accountable. In the 1990s, Congress became involved in standards-based reforms through *Goals 2000*, which sparked more federal government involvement via Title 1 funds (Superfine, 2011). The *Goals 2000* legislation codified the National Education Goals and offered grants to states that committed to specific plans for systematic reform of K-12 education, which included standardized testing of reading and mathematics skills to ensure students' performance of academic standards. The federal government was lenient in holding states accountable due to opposition between political parties. However, this marked the beginning of the federal government using federal money to leverage systematic change. McGuinn (2015) states:

> National policymakers have used the grant-in-aid system to pursue federal goals in public education. To claim their share of the growing pot of federal education funds, states had to agree to comply with a wide array of federal policy mandates. (p. 77)

Beginning in the 2000s, the federal government's role in education changed dramatically through unprecedented reform efforts leveraging federal funding for systematic change. This has been accelerated by the reauthorization of the *Elementary and Secondary Act* of 1965 called the *No Child Left* Behind of 2001 (NCLB), *A Blueprint for Reform: The Reauthorization of the Elementary and Secondary Education Act of 2010,* and *Every Student Succeeds Act of 2015* (ESSA) (United States Department of Education, 2004; United States Department of Education, 2010; United States Department of Education, 2015). Although the original passage of ESEA in 1965 focused on distributing federal funds according to the number of low-income students

enrolled in schools, increasing student achievement and accountability measures through leveraging federal funds have been at the forefront of educational reform ever since. Federal funding increased from $1.5 million to approximately $60 billion between 1965 and 2010 (McGuinn, 2015). With each reauthorization of the ESEA, the federal government has assumed a much greater role in funding, authority, and accountability measures.

Under NCLB, states were forced to set academic standards that moved from norm-based testing to criterion-based testing, which focused on achievement levels over time through high-stakes testing and gave the government unprecedented authority when schools failed to meet NCLB's performance targets (United States Department of Education, 2004). This mandated state annual summative tests in grades three through eight in reading and mathematics to claim Annual Yearly Progress (AYP) on measurable educational objectives (McGuinn, 2015). Federal government involvement through NCLB's AYP enabled school monitoring via student test scores and set a new precedence for additional services and school restructuring. This included the implementation of highly qualified teachers in every public-school classroom to ensure students were 100% proficient by 2014 deemed by student test scores. McGuinn (2015) states:

> One of the most important mandates in the law is that school report cards must disaggregate student test score data for subgroups based on race and ethnicity, economically disadvantaged status, limited English proficiency, and classification as in need of special education. (p. 86)

Persistent failure to meet AYP performance targets by any subgroups required states to take interventive action in school districts and schools. States were forced to create new testing and accountability systems, which elevated and expanded the states' capacity for involvement in monitoring school districts and schools. The NCLB aggressively expanded the federal

government's role and authority in education by forcing states to create accountability and assessment systems to report student scores based on students' annual summative test scores.

The Obama administration released its, *A Blueprint for Educational Reform*, the reauthorization of ESEA in 2010, which built upon the core principles of NCLB and held states to rigorous standards by developing and implementing college and career-ready academic standards with aligned assessments and required states to provide intensive support and interventions (United States Department of Education, 2010). Additionally, the *American Recovery and Reinvestment Act* (ARRA) was put into place to stimulate and stabilize the economy during one of the worst financial crises since the Great Depression. The ARRA was a call to action for educational reform through federal government initiatives. Many education reform efforts were built on the ARRA with a focus on managing the teacher workforce, balancing local and federal issues, and drawing on educational research for a more effective federal action in education. According to Superfine (2011):

> The ARRA was designed to keep school systems afloat in difficult times and it is also aimed at fixing existing educational policy problems and sparking future educational reform efforts. The ARRA framed the Obama administration's subsequent educational reform efforts including the impending reauthorization of the No Child Left Behind Act of 2001. (p. 2)

The *Race to the Top* (RTT) initiative sought to continue the reform efforts of the Obama administration to ensure all students were college and career ready through a grant competition to selected states that "agreed to implement a range of education policies and practices designed to improve student outcomes" (U.S. Department of Education, 2016, p. 5). This constituted even more nationwide changes as RTT placed great emphasis on the design and implementation of

rigorous standards and high-quality rigorous assessments, which placed more emphasis on the results of student standardized test scores. The U.S. Department of Education used RTT as a strategy for leveraging federal funds by getting the states to agree to their priorities (Tatter, 2015), which increased the power and authority of the federal government in public education tremendously. Kornhaber et al. (2016) suggest the federal government used competition for critical resources during a great recession that enabled extraordinary influence over most states' academic standards. Superfine (2011) states, "The federal government should use its considerable power to leverage educational reform strategies that at least have substantial evidence underlying claims about their efficacy instead of just political consensus" (p. 15).

Out of the ARRA $65 billion educational reform effort, Tennessee received over $500 million of RTT funds. This created an unprecedented and intensive focus on adopting new rigorous academic standards, assessments, data systems to measure student growth and success via student test scores (Tennessee State Government, 2010b). Additionally, the RTT targeted the lowest-performing schools and highlighted the importance of teacher development to improve instruction via designing policies to reward and retain teachers. To help ensure Tennessee's competitiveness for RTT funds, Tennessee passed the *First to the Top Act of 2010* (First to the Top) (Tennessee State Government, 2010a). The First to the Top law encapsulated most of the RTT provisions such as adopting the Common Core State Standards (CCSS) for math and English and creating a new evaluation system for Tennessee teachers, which included student test scores on state-mandated summative tests (Tatter, 2015). Half of the RTT funds were distributed to school districts while the other $250 million was used for statewide systematic reform to revolutionize Tennessee education.

The First to the Top law established the Tennessee Achievement School District (ASD), which allowed state intervention in the lowest-performing schools, and established new requirements for annual teacher evaluations that incorporated student test scores from the end-of-the-year state summative tests (Tennessee State Government, 2010a). Additionally, certain student-achievement data would be used in making decisions about teacher tenure. The First to the Top mandated 50 % of teacher evaluation ratings be based on student testing data from state standardized tests (Tatter, 2015). Thirty-five percent is based on TVAAS (Tennessee Value-Added Assessment System) or other comparable measures of student growth for teachers in non-testing grades and 15 % is to be determined by teachers and evaluators on other testing data. TVAAS is a measurement system used to measure student academic growth on state summative standardized tests as evidence of teachers' impact on student learning from one year to the next. TVAAS scores are used to reward or penalize teachers. Before this law, TVAAS was used for diagnostic purposes only and did not come with consequences for educators.

According to Tatter (2015), teachers appeared to get a double message from leaders. On one hand, teachers were told don't teach to the test, yet on the other hand, teacher evaluations are based on student state summative test scores. This marked another significant increase in federal and state influence and authority in Tennessee public education and a significant focus on student standardized test scores for educators in Tennessee. This mobilized school districts to use practice tests to ensure student preparedness and performance on state summative tests and remains a focus for Tennessee educators.

Lastly, under the Obama administration, the reauthorization of the ESEA called the *Every Student Succeeds Act* (ESSA) was signed into law in 2015 (United States Department of Education, 2015). The ESSA has attempted to release some of the federal's government power

over testing and evaluation measures back to the states. States are responsible for adopting challenging academic achievement standards that must be aligned with relevant state career and technical education standards.

For instance, Tennessee adopted the CCSS in 2010 with the implementation in the 2011-2012 school year (Balakit, 2016). Due to public and political pushback against CCSS as government overreach, Tennessee phased out CCSS with the creation and implementation of Tennessee Academic Standards for math and English language arts in the 2017-2018 school year. However, TVAAS and the teacher evaluation system that relies on student scores on state summative tests remain in effect and at the forefront of testing and accountability measurements for Tennessee educators.

Current Government Role in Education Amid the Covid-19 Pandemic

Since the onset of the Covid-19 Pandemic in 2020 the United States Congress has passed three stimulus spending packages with specified funding for elementary and secondary schools (Bernstein, 2021), which surpass all previous federal government spending allocations. In March 2020 the *Coronavirus Aid, Relief, and Economic Security* (CARES) *Act* was passed with $13.5 billion for elementary and secondary school emergency relief, which created the ESSER fund. The ESSER 2.0 fund totaling $54.3 billion was established through the passage of the *Coronavirus Response and Relief Supplemental Appropriations Act* (CRRSA). And the most recent ESSER 3.0 fund totaling just under $123 billion was passed under the *American Rescue Plan* (ARP) act in March 2021.

The ESSER funds came with plenty of strings attached, which subsequently increased with each legislative package and included when and what the funds could be spent on in schools. The CARES Act stipulates the funds are for financial support to school districts and

schools dedicated to safety and healthy return to school and continuing equitable education opportunities and services for students (Bernstein, 2021). However, the federal government expanded guidance and requirements for ESSER 2.0 and ESSER 3.0 funds. With ESSER 2.0 funds, the uses included preparing schools for reopening, upgrading air quality in schools, and addressing learning loss. With ESSER 3.0 funds, the federal government requires a minimum of 20% to be allocated to address learning loss and related challenges due to Covid-19 interrupting American education. The remaining 80% of ESSER 3.0 follows the same guidelines as ESSER 1.0 and ESSER 2.0.

Tennessee received almost $4.2 billion in total ESSER funds. $260 million in ESSER 1.0, $1.1 billion in ESSER 2.0, and nearly $2.5 billion in ESSER 3.0 (Tennessee Department of Education, 2021c). ESSER funds must be spent between spring 2020 and fall 2024. The distribution of ESSER funds follows the Title 1 formula established under the ESEA to the states and then to school districts within the states (Bernstein, 2021). The states must distribute 90% of ESSER funds to local education agencies (LEAs), thus infusing and levering state government purse strings towards LEAs.

Tennessee's strategic spending plan for ESSER funds is infused with key pieces of Tennessee State Legislation to address education issues resulting from the Covid-19 pandemic disruptions such as instructional learning environments and student learning loss (Tennessee Department of Education, n.d.a). The *Tennessee Learning Loss And Student Acceleration Act (SB7002/HB7004)* and the *Tennessee Literacy Success Act (SB7003/HB7002)* became new education laws in Tennessee in January 2021. This positioned the Tennessee Department of Education to support its Reading 360 initiative, which proposes to tackle the literacy crisis in Tennessee, through legislative action and the use of federal funding streams.

The *Tennessee Learning Loss And Student Acceleration Act* (Learning Loss Act) established two summers of learning loss and remediation camps for Grades 1-8 that began in the summer of 2021. Guidelines were set for prioritizing students' participation, referred to as priority students, and camp requirements of daily instruction as four hours of math and literacy instruction and additional time for RTI^2 instruction (Tennessee Department of Education, n.d.c). Moreover, this turns into a permanent bridge program for students entering Grades 4-8 who need additional learning time after ESSER funds are gone. Additionally, this law established the *Tennessee Accelerating Literacy and Learning Corps* (ALL Corps) to ensure the availability of qualified tutors This law also established a 3rd-grade reading gate, which provides third-grade students additional time and academic support through the ALL Corps during their 4th grade school year based on student state summative test scores.

The second piece of legislation was the *Tennessee Literacy Success Act* (Literacy Success Act) law. The Tennessee Department of Education (n.d.d), states this legislative law, "lays a policy foundation for literacy in Tennessee to ensure every STUDENT builds strong reading skills grounded in phonics, and every EDUCATOR is supported with training and resources to become a strong literacy instructor" (p.1), which established:

- Every district must use phonics-based literacy skills to teach reading.
- Require educator preparation providers to provide training on foundational literacy skills.
- District Foundational Literacy Skills Plans to boost literacy for all students.
- Free Tennessee universal reading screener option to gauge student progress.
- Required foundational skills training for teacher candidates in educator preparation programs.

- Required reporting to the legislature and public on statewide literacy progress. (Tennessee Department of Education, n.d.d, p.1)

Additionally, the Literacy Success Act requires foundational training for all K-5 teachers via the state department, which consists of two courses. At the end of each foundational training, Tennessee educators must pass the state department's created end-of-course tests.

The Tennessee Department of Education (n.d.d) used student state standardized testing data and third-grade reading proficiency research as a foundational cornerstone for this law stating:

- Before the COVID-19 pandemic, only one-third of Tennessee third graders and 27 percent of eighth-graders were proficient in English Language Arts.
- Data from the Nation's Report Card shows that the reading proficiency of fourth-graders in Tennessee has remained relatively unchanged for the past 7 years.
- If students cannot read on grade level by third grade, it is much harder for them to catch up with their peers and to succeed in their education.
- Children who are reading on grade level by third grade are significantly more likely to stay on grade level over time; graduate from high school, enter and complete post-secondary programs; and become gainfully employed later in life. (p.1)

Due to the continuation of federal and state legislative requirements placing significant importance on high-stakes testing as the primary method of measuring educational success and progress and the associated streams of school funding from government entities, school district leaders and teachers continue to realize the need and urgency to measure student progress throughout the school year. This creates a reliance on using other forms of assessments effectively, such as formative computerized adaptive assessments, (Bruce, 2019; Thompson,

2018), to ensure student progress and achievement in mastering skills and knowledge assessed by state academic standards (Stiggins, 2005). This places an enormous challenge on educators to find and utilize formative assessments for predictive purposes that will provide adequate and accurate data of student performance levels and progress throughout the school year as measured on the state summative standardized tests.

History of Formative Assessments

Due to the emphasis on accountability, student achievement and student improvement have become a motivating force for educators as they make decisions about instruction. Educational systems are continuously seeking avenues such as formative assessment systems to provide data to increase student learning, achievement, and improvement throughout the teaching and learning cycle as measured by student test scores on high stakes annual tests. Formative pedagogy practices and assessments are an integral part of many of today's educational institutions.

Formative Assessments

The relationship between teaching and learning is complex with many intertwined facets. Educators use formative assessments to measure the effectiveness of the teaching and learning cycle to increase student learning and achievement more intensely due to modern accountability systems, which are based on state standardized test scores. Using formative assessments as an effective means of measurement within the teaching and learning cycle was spurred greatly by two articles by British researchers Paul Black and Dylan Wiliam. The first article by Black and Wiliam (1998a) was a seminal review of literature of more than 250 empirical research studies focused on the effects of classroom pedagogy and assessment practices concerning students' increased learning and achievement outcomes. In the second article by Black and Wiliam

(1998b), the researchers took a firm stance that formative assessments data must be used to "adapt the teaching to meet student needs" (p. 140). The researchers supported their stance from evidence from their meta-analysis review, in which they concluded: "Formative assessment does improve learning. The gains in achievement appear to be quite considerable, and as noted, amongst the largest ever reported" (Black & Wiliam, 1998a, p. 61).

Black and Wiliam (1998b) made several conclusions that remain relevant as leadership teams seek to optimize school improvement efforts through proper utilization of formative assessments within the instructional design and implementation process. Black and Wiliam (1998b) stated that for an assessment to be formative, the information collected must be used to modify instruction to meet student needs. Testing for testing's sake and to accumulate data does not lead to increased student achievement. Successful practices that involved dramatic changes in pedagogy involve active student engagement in the assessment process. The researchers urged the relationship between feedback provided during the formative assessment process and the student effect. The methods for feedback communications to students needed to produce different outcomes to encompass student ownership of their learning through a lens of positivity. Black and Wiliam (1998b) concluded through research evidence that the use of formative assessments in classrooms can raise achievement, but "only by changes that are put into direct effect by teachers and pupils in the classrooms" (p. 148). Testing, with the purpose of improvement, within the teaching and learning cycle is paramount to raising student learning and achievement.

Defining Formative Assessments. The insurgent interest in formative assessments being touted as 'best practice' for increasing student learning and achievement led to divergent definitions and uses for formative assessments. "Formative assessment's status as an ethereal

construct has further been perpetuated in the literature due to the lack of an agreed-upon definition" (Dunn & Mulvenon, 2009, p. 2). Many educational leaders and teachers are eagerly investigating resources in leveraging formative assessments to raise student performance on high-stakes testing; therefore, they need to have a common understanding of what constitutes formative assessment. The Council of Chief State School Officers (CCSSO) defined formative assessment by the Formative Assessment for Student and Teachers (FAST) as a process used during instruction to provide feedback to adjust or make modifications for ongoing teaching and learning to improve student achievement as related to instructional objectives (Popham, 2006a). Popham (2008, as cited Dunn & Lulvenon, 2009) expanded his definition of formative assessment as a planned process that the teacher and student use the assessment data as evidence to modify or adjust ongoing instruction and learning. According to Filsecker and Kerres (2012), formative assessments are a series of informed and informing actions that change the current state of knowledge to a more knowledgeable one within the teaching and learning cycle. Formative assessments are short-cycle assessments providing a timely turn-around for data use by educators to inform instructional decisions, planning, and implementation of instruction (Perie et al., 2007; Popham, 2006b) in context with the objectives and standards that have been taught to gauge student learning.

Formative Assessment Purpose. Educators must have a clear vision of how formative assessment systems can increase student learning and achievement in the teaching and learning cycle within the school improvement process for all stakeholders. According to Stiggins (2002), "If we are finally to connect assessment to school improvement in meaningful ways, we must come to see assessment through new eyes" (p. 758). With formative assessments, teachers and students contribute to a continuous flow of information related to student achievement. To

enhance and advance the benefits of formative assessments for learning, Stiggins (2002) lists the following components: from the beginning students need to recognize and understand the expected achievement goals and targets, teachers must use descriptive feedback to build on prior knowledge, the teacher must continuously adjust instruction to meet the needs of the students, and students must regularly engage in self-assessment.

Dunn and Mulvenon (2009) noted that assessments can serve a variety of roles in education including diagnosing weaknesses, evaluating student understanding within the teaching and learning cycle, and predicting student performance on summative assessments, which is at the forefront of school improvement efforts and initiatives by leaders and teachers seeking to find effective tools to provide data-based evidence that will lead to increased student learning and achievement. To arrive at the desired destination of student learning and achievement for all students and to meet the demands of high stakes testing through mandated accountability structures set forth by policymakers, educational leaders must have a clear vision of where they want to go, how they plan to get there, and by what vehicle will transcend them to their destination. Formative assessments provide educators with data-based evidence, which can provide data to identify learning gaps to inform instruction and empower and motivate students to increase student learning and achievement regarding instructional objectives and standards through an ongoing feedback loop within the teaching and learning cycle (Popham, 2006a; Stiggins, 2005).

Interim Assessments

According to Marshall (2008), interim assessments encompass three powerful insights. First, initial teaching cannot bring all students to proficiency level due to the uniqueness each student brings to the teaching and learning cycle. Second, educators should not wait until the end

of the year exams to determine student learning. And third, with much dedication, educators can address and close learning gaps before they snowball. Additionally, Marshall (2008) states many good teachers have applied these insights to bring out the best in students, which is confirmed by three strands of research Benjamin Bloom's work on mastery learning, the "effective schools" research, and Total Quality Management.

Defining Interim Assessments. Benchmark, diagnostic, formative, and or predictive assessments provide data for improving student performance levels within the school year and fall under the umbrella term 'interim assessment' (Perie et al., 2009; Perie et al., 2007). Interim assessments are medium-scale, medium-cycle assessments occurring between summative and formative assessments, which are usually administered at the district or school level (Perie et al., 2009). Perie et al. (2009) define interim assessments as:

> Assessments administered during instruction to evaluate students' knowledge and skills relative to a specific set of academic goals in order to inform policymaker or educator decisions at the classroom, school, or district level. The specific interim assessment designs are driven by the purposes and intended uses, but the results of any interim assessment must be reported in a manner allowing aggregation across students, occasions, or concepts. (p. 6)

Perie et al. (2009) state the definition of interim assessments has two key components "(1) evaluate students' knowledge and skills relative to a specific set of academic goals, typically within a limited time frame, and (2) are designed to inform decisions both at the classroom and beyond the classroom level, such as the school or district level" (pp. 6-7). Marshall (2008) suggests interim assessments check for understanding within several weeks of the teaching and learning cycle. Additionally, interim assessment data utilized in a more formalized assessment

system can provide an additional lift that can take teaching and learning to even higher levels. When interim assessments are done correctly, Marshall (2008) states, they can have a ripple effect that can fuel improvements in every other stage in the teaching and learning cycle by helping teachers plan better, teach better, and use interim assessment data more powerfully to close the achievement gap during the school year.

Interim Assessment Purpose. Interim assessments can be categorized for serving three general classes of purposes within the interim assessment system: instructional, evaluative, and predictive (Perie et al., 2009; Perie et al., 2007). According to Perie et al. (2009), the primary goal of an interim assessment for instructional purposes is to use the data to adjust instruction and or curriculum to accommodate student learning needs. Interim assessments that fall under this category can be used to identify students' strengths and weaknesses in their learning related to content, skills, and standards; to modify instruction; to provide an in-depth look at student misconceptions; and, lead to strategies for improving instruction. The researchers suggest for interim assessments to serve an instructional purpose, the assessment system must provide educators with strategies for interpreting and using data to modify instruction effectively in the classroom. Therefore, this type of interim assessment could also be tagged formative.

The second purpose an interim assessment system can serve is evaluative to provide information about curriculum and instruction (Perie et al., 2009; Perie et al., 2007). The researchers define an evaluative assessment design as one that would explicitly provide data to help teachers, principals, district personnel, and decision-makers to learn more about the curriculum or instructional choices and take specific actions for improvement affecting teaching with the intent to improve learning. Lastly, the researchers contend evaluative assessment systems must provide detailed data about curricula.

The third purpose of an interim assessment system design can serve is for predictive purposes, which is to determine each students' likelihood of meeting performance expectations on end-of-the-year exams (Perie et al., 2009; Perie et al., 2007) such as on high-stakes summative tests. The researchers contend many users seek informational data to assist them in improving the performance levels of students for whom failure or low performance is predicted or to aid in the prediction of low performance on future tests. Additionally, the information may derive from the assessment itself or from further probes to identify areas of weaknesses of those not on track for success as indicated by the interim assessment. In this scenario, interim and formative assessments could work together to improve student performance levels on summative assessments, which highlights the importance of aligning all facets of a comprehensive assessment system. A quality interim assessment can be an indispensable part of a comprehensive assessment system when used in conjunction with classroom formative and summative state assessments at the district or state level.

Additional Considerations

Perie et al. (2009) noted professional learning needs to accompany any assessment system and the ownership of responsibility should be from the vendor or the education system. This is a source of contention for many assessment developers to provide professional learning. Additionally, Marshall (2008) states traditionally many principals and educational leaders evaluate and supervise the process of teaching and learning rather than discussing the results. When utilizing formative assessment data, the challenge for education leaders is having ongoing conversations with teachers to effectively use testing data within the teaching and learning cycle. These conversations include getting teachers to slow down, reflect on what is working and what is not, and integrate a continuous cycle of improvement driven by insight from real-time

assessments. These conversations are vital due to changes in teachers' practices being deeper and more lasting when they come from and through collaboration about the best ways to get all students to higher levels of achievement, which includes using and incorporating formative assessment data effectively and efficiently. This also could help reduce teacher attrition when considering the current strenuous accountability measures placed on teachers with an emphasis on student test scores.

Teacher retention is at its worst in those first, early years. On average, 44% of teachers will leave within the first five years alone (Ingersoll et al., 2018), with some schools faring even worse. Principals and education leaders must take the lead role in retaining teachers to maintain a viable workforce with reducing teacher attrition being one of the principal's most challenging tasks (Norton, 2015). A report by the Consortium for Policy Research in Education revealed that our national teaching force is unstable and new teachers continue to be among the highest attrition rates (Ingersoll et al., 2018). Additionally, the teaching force is getting older, and the report predicts a dramatic increase in the demand for new teachers (Ingersoll, 2012; Ingersoll et al., 2018).

When considering why so many teachers choose to leave the profession, research reveals job dissatisfaction across a variety of school and working conditions, which includes lack of leadership support, salaries, classroom resources, student discipline, opportunities for growth and development, and lack of input into decision making within the school (Ingersoll, 2001). According to Norton (2015), "Lack of support from administrators outranked eleven other aspects of the job by a large margin as a factor for teachers leaving or considering leaving their jobs" (p. 52). Teaching requires intense interactions with children; yet, the work of teachers is done largely in isolation, which can be difficult for new teachers as they are by in large left to

succeed or fail within their classroom (Ingersoll, 2012). Induction supports have a positive effect on teacher attrition (Ingersoll, 2001, 2012; Ingresoll & Smith, 2004). According to a study by Ronfeldt and McQueen (2017), which used the most recent data from the Schools and Staffing, Teacher Follow-Up Surveys, and the Beginning Teacher Longitudinal Study, new teachers receiving induction supports reduce new teacher turnover. New teachers may feel ill-equipped or less confident in aspects of day-to-day job tasks and expectations such as effectively using formative assessment data.

When considering national and state legislation pressuring school districts to raise state summative test scores, this places an enormous challenge on educators to find and utilize formative assessments that will provide adequate and accurate data on student performance and progress toward meeting state academic standards. Educators, especially those in leadership, must understand the rationale behind formative assessment research to provide a coherent and highly functional assessment system. For assessment systems to be truly effective at the district and school level, they must be able to provide different kinds of information to various policy and decision-makers at different times in various forms for three different levels the classroom, the school, and the district (Stiggins & DeFour, 2009). Education leaders must provide effective assessment practices and systems in the teaching and learning cycle to ensure student learning and achievement and provide teacher support for effective data use.

Assessments for Accountability Measures in Tennessee (Standardized Tests)

Although assessments have been a part of the educational system for more than 100 years, educational assessment methods and formats have changed over time (Shepard, 2016). The overall purpose of using assessments to measure student achievement, evaluate student progress, and inform instruction is a common well-established practice in education and remains

at the forefront of educators' decision-making. High-stakes tests are a method of assessment that students, teachers, school districts, and states use to measure a learner's ability and enforce academic accountability "In general, "high stakes" means that test scores are used to determine punishments, accolades, advancement, or compensation" (Great Schools Partnership, 2014, para. 1). High-stakes tests are often referred to as summative tests, state standardized summative assessments or tests, and standardized tests. Perie et al. (2007) state summative assessments are typically given at the end of the semester or school year "to evaluate students' performance against a defined set of content standards" … and "are usually part of an accountability program or to otherwise inform policy" (p. 1). Summative standardized tests are typically not flexible in the design or testing procedures. Standardized testing is the primary form of student learning assessment facilitated by current accountability measures and systems (Styron & Styron, 2012), which typically seeks to determine students' understanding, knowledge base, and skill level of current grade-level academic standards in preparation for the next grade level academic standards. Data utilized from summative standardized test scores can impact decisions at every level in a school district, such as grade retention, student graduation, school curriculum, teacher evaluations, funding, and government interventions.

Tennessee Department of Education (TDOE) requires end-of-the-year summative assessments for Grades 3-8 with 2nd grade being optional as part of the state's Tennessee Comprehensive Assessment Program (TCAP) (Tennessee Department of Education, n.d.e). The purpose of TCAP assessments is to assess students' skill level and understanding of the Tennessee Academic Standards and provide a summative measure of achievement in reading, math, science, and social studies as well as provide reports on student performance levels to school districts, teachers, and parents. Additionally, these summative assessments are

administered in compliance with the *Every Student Succeeds Act of* 2015 and T.C.A. § 49-1-602 for district and school accountability.

TCAP has been Tennessee's testing program since 1988 and includes TNReady assessments (Tennessee Department of Education, (n.d.e). Tennessee began using TNReady assessments, which are aligned with Tennessee's more rigorous academic standards and surpass ESSA requirements, in 2015-2016 (Tennessee Department of Education, 2018). The creation of TNReady assessments consists of a three-step collaborative evaluation process (Tennessee Department of Education, n.d.b). The first step consists of a collaboration between the state department, Tennessee teachers, and testing vendors to create test questions. The state department sets the expectations for test questions in alignment with the Tennessee Academic Standards, whereas teachers and test vendors develop test questions. The second step consists of a collaboration between state content teams and vendors to ensure alignment and standard expectations. Tennessee teachers' committees are to ensure content alignment and no bias issues relative toward specific student groups or sensitive topics exist, and vendors are to make necessary updates based on the teachers' feedback. The final step consists of the state department and vendors putting the tests together. Each test question is field-tested and reviewed for statistical validity. After test questions are deemed statistically valid, the vendor adds the questions to an operational test followed by a review by the state department for accuracy. Finally, the final TNReady assessment is administered in Tennessee schools.

Each student, school, and school district receives TNReady score reports (Tennessee Department of Education, n.d.f). TNReady score reports provide detailed information about students' achievement performance in comparison to the subject and grade-level expectations for the Tennessee Academic Standards. The score reports include scale scores and percentage of

student performance level, which is based on the following performance levels: mastered, on track, approaching, or below. Lastly, the score reports a comparison of student performance levels with other students in their school, district, and state.

Before 2020, Tennessee had risen from being one of the lowest-ranked states in reading and math proficiency to around the national average on the National Assessment for Educational Progress (NAEP) in 2017 but a decade of previous gains has practically disappeared on the 2020-2021 TCAP data due to Covid-19 learning disruptions (State Collaboration on Reforming Education, n.d.). The TNReady data for the 2020-2021 school year showed a decrease in student achievement across all subjects and grades as a result of pandemic-related disruptions in education (Tennessee Department of Education, 2021d). An overall view of state-level TCAP results specifically found in Grades 3-8 and optional Grade 2 assessment:

> English Language Arts proficiency dropped 5 points from 2019.
>
> 3 in 10 Tennessee students are meeting grade-level expectations in ELA.
>
> 2nd & 3rd grades scores showed large increases to students scoring Below.
>
> 1 in 4 Tennessee students is on grade level in math.
>
> Overall 3rd-grade proficiency declined from 44% in 2019 to 31% in 2021, while 4th-grade proficiency declined from 46% in 2019 to 34% in 2021. (Tennessee Department of Education, 2021d, p.1)

The state department acknowledged the challenges of the Covid-19 pandemic such as the mode of instruction and scheduling and the results being distinctively different from previous years. However, these results show the importance and urgency of addressing student needs due to Covid-19 pandemic-related student learning loss for K-12 education. According to the Tennessee Department of Education (2021d), the state department recognized the impending

impact the Covid-19 would have on education in Tennessee and has taken a proactive and strategic approach utilizing portions of the state's ESSER fund spending plan. The goal is to benefit K-12 education to address and combat student achievement gaps, the needs of rural communities, and to accelerate student academic achievement through legislative action, the Learning Loss Act and the Literacy Success Act, on accountability, learning loss, and literacy deemed via TCAP standardized testing data.

The Tennessee Public Education Coalition (2021) suggests the reported student data, which was derived from state summative test scores used before the Tennessee General Assembly to push the Literacy Success Act through legislation, was misleading statistics. For example, most legislators who likely heard the statistic did some quick math and assumed approximately 63% of third-graders and 73% of eighth-graders are reading below grade level. Whereas, the state's 2019 TNReady data less than 20% of students are reading below grade level in grades three through eight, and the proficiency percentages of 45.5% of students in the 'approaching' category in reading were not included. Students in the 'approaching' category are not below grade level.

Additionally, The Tennessee Public Education Coalition (2021) suggests the use of misleading data was used in passing the Learning Loss Act, which would require schools to retain an expected 60% of third-graders starting in 2023 and could be damaging to students being held back. According to The Tennessee Public Education Coalition (2021), using misleading student data statistics "created a misconception of a crisis to push through questionable policy and programs that will enrich preferred education vendors" (para. 1). Consequently, Bulkley and Burch (2011) suggest education vendors' role and revenue continue to increase in public schools due to policy mandates by expanding services. For example, following the NCLB passage, top

education vendors' revenues increased from $430 million in 2003 to $780 million in 2008. Additionally, Bulkey and Burch (2011) report education vendors also influence many critical aspects of public-school governance. As a result of accountability reforms, many school districts pay outside vendors to design and assist in a wide array of interlocking services and products. Many vendors offer school districts a one-stop-shopping approach for services and products. Bulkey and Burch (2011) state the hiring of private firms is an extension of educational policy processes, which sets “preferences for what educational outcomes matter, to track educational outcomes, and to design interventions based on these outcomes” (p. 241).

However, The State Collaboration on Reforming Education (SCORE) (n.d.) strongly supports accelerating student learning through high doses of tutoring and summer learning to offset the interruptions of student learning and achievement due to overcoming the disruptions of learning caused by the Covid-19 pandemic. Yet, as a result of this law, beginning in the summer of 2021 thousands of Tennessee students could be required to attend summer school for six weeks, five days a week, for seven and half hours per day based on their state test scores (The Tennessee Public Education Coalition, 2021). Conversely, SCORE (n.d.) suggests the results from the TCAP are a reliable indicator of student achievement. Student state summative test scores continue to be at the center of Tennessee policy mandates for public education in Tennessee.

There are several new built-in assessment and accountability measures in the Learning Loss Act based on student TCAP data and or state-approved assessments (Tennessee Department of Education, 2021b). All LEAs are required to provide all priority students with the opportunity to enroll in summer camps and local school boards can adopt policies requiring priority student attendance. “Priority students include students who score below proficient in reading or math on

the most recent state test or state-approved screener or students who attend a school where less than 50% of students are proficient in math or reading" (Tennessee Department of Education, 2021b, p.2). School districts must also administer state-approved pre-and post-test assessments during the summer camps and report student testing data to the state department. Beginning in the 2022-2023 school year, third-grade students scoring at the "approaching" or "below" performance level on the ELA TCAP, state-mandated summative tests, are required to receive additional supports and services, which may include summer camps.

Accordingly, there are several new assessments and accountability measures built in the Literacy Success Act regarding student test scores that are interrelated across several continuums for Tennessee school districts and teachers. State-approved universal diagnostic screeners must be administered to all K-8 students three times a year, which also meets the requirements for RTI^2 and the *Tennessee Dyslexia Law (T.C.A. § 49-1-229)* (Say Dyslexia Law) to identify students who may need extra support and or other types of instruction (Tennessee Department of Tennessee of Education (2021a). Teachers of Record, normally meaning classroom teachers, are not allowed to administer universal reading screeners unless they are a web-based screener, in which classroom teachers are allowed to monitor students taking the assessment but not allowed to enter student scores for state department use. Additionally, teacher of record information including teacher license numbers must accompany student results to the state department after each of the three universal screeners, which identify teachers and students respectively. Lastly, home literacy reports must be sent for students identified as at risk for a significant reading deficiency, which is a composite score at or below the 15^{th} percentile. However, the state department encourages home literacy reports to be sent home with students who score between the 16^{th} and 40^{th} percentile. Student assessment scores on state-approved assessments and

accountability measures remain at the forefront for Tennessee educators across a wide array of continuums layering an intensive focus on student testing data for school districts, schools, and teachers and influencing the use of formative computerized adaptive tests to predict student performance level and success on state standardized tests as well as diagnostic assessments and screeners.

Formative Computerized Adaptive Tests

Computerized adaptive tests (CAT) are assessments that adjust to the level of knowledge of the student based on the item response theory (IRT). IRT is based on a statistical method used to predict a student's ability to achieve a score based on unobservable latent traits such as aptitude, achievement, skill, and personality and their manifestation such as outcomes, responses, or performance (Zanon et al., 2016). IRT aims to establish a relationship link between tested items, individual responses to those tested items, and the underlying latent traits being measured (Lang & Tay, 2020). CAT is structured on learning progressions based on sequential skill competencies across grade levels as defined by academic content standards (Bennett, 2015; Shapiro et al., 2015). The goal is to pinpoint students' achievement location along the learning progress continuum at a given time with validity, reliability, and fairness to all demographics (Bennett, 2015) to provide evidence for data-based decision-making processes.

When using CAT in this manner, they essentially become formative computerized adaptive tests or assessments providing student data for education decision-makers to use constructively within the teaching and learning cycle serving three broad purposes: instructional, evaluative, and predictive purposes (Immekus & Atitya, 2016). To further support the action of computerized adaptive assessments functioning formatively, the use of an assessment and inferences made regarding the functionality of the student data determines the assessment's

classification as summative or formative (Wiliam, 2014). Inferences made from data regarding students' current or future learning and achievement are derived from the summative assessments, whereas inferences made regarding the supports likely to improve student learning and achievement are derived from formative assessments. However, “an assessment functions formatively to the extent that evidence about student achievement is elicited, interpreted, and used” (Wiliam, 2014, p. 5) by educators for decision-making purposes. Furthermore, the functionality of information and data gleaned from an assessment serves as the distinction of use, not on the kind of assessment. Summative and formative assessment purposes are not mutually exclusive but can coexist (Van der Kleij et al., 2015). Choi and McClenen (2020) propose a computerized adaptive formative assessment system (CAFT) that can be used to administer a personalized adaptive assessment for which students and teachers receive personalized diagnostic feedback about students’ learning progressions with multiple testing points within a course of study. Additionally, the researchers state there has been limited research on online formative assessments functionality based on statistical methods.

Therefore, when viewed from these perspectives, formative computerized adaptive assessments and assessment systems, which include interim, benchmark, and diagnostic computerized adaptive assessments (Perie et al., 2009; Perie et al., 2007), can serve as an approach and a decision-making vehicle to provide educators access to evidence of student learning, achievement, mastery of skills, and for predictive purposes via student test scores (Bruce, 2019; Immekus & Atitya; 2016; Thomas, 2018) as well as provide evidence of an assessment system’s effectiveness (Stiggins & DeFour, 2009). This conceptual basis is grounded in the integration of data-based decision-making, in assessing for learning and of learning, through online formative and diagnostic testing approaches (Bruce, 2019; Collares & Cecilio-

Fernandes, 2019; Choi & McClenen, 2020; Immekus & Atitya, 2016; Perie et al., 2009; Perie et al., 2007; Sutter et al., 2020; Thomas, 2018).

Benefits of formative computerized adaptive assessments include real-time updates that are provided on a student's learning progression within a designated timeframe (Choi & McClenen, 2020; Collares & Cecilio-Fernandes, 2019). Such information can offer quick remedial actions and feedback to students, teachers, and leadership teams. The success of computerized formative adaptive assessments is fundamentally grounded on accurate information about what a student has learned and knows about specific skills and concepts related to academic standards (Choi & McClenen, 2020). As such, a valid and meaningful formative computerized adaptive assessment system provides an accurate assessment of a student's ability and diagnoses what skills, concepts, and standards a student knows and has learned. Improved efficiency allows for the precision of measurement and reduces testing time with more accuracy than conventional tests (Center on Standards and Assessments Implementation (CSAI), 2019). In using formative computerized adaptive assessments, past performances can be used as a starting point to measure student growth as well as to pinpoint areas of weakness. This also provides more test security due to the matching of test items to students' results and reduces the chance and risk of cheating. However, limitations of using CAT programs may include a lack of school infrastructure, a sufficient number of student devices for successful implementation, students cannot skip questions as on conventional tests, the size of the item test bank if it is not sufficient to cover a large range of skills, up-to-date content with skill alignment (CSAI, 2019), and algorithm issues on test question differentiation based on student responses (Thompson & Weiss, 2011).

There is minimal research available on the predictive use of student test scores from formative computerized adaptive assessments to guide decision-makers at the classroom, school, and district levels (Immekus & Atitya, 2016; Sutter et al., 2020). A key component of using formative computerized adaptive assessment data is the ability to examine the relationship of student scores to state summative standardized test performance. Using student test scores from formative computerized adaptive assessments can provide information for effective decision-making for teachers at the classroom level and school and district level for evaluative and predictive purposes, such as for the product, program, or assessment system's effectiveness.

In response, the researchers investigated the technical quality and validity of English Language Arts (ELA) district-developed computerized adaptive interim assessment test data to predict student performance levels on state summative tests (Immekus & Atitya, 2016). This study was based on Grade 6 student data (N = 4,651) from a large, urban school district, which included demographics, scale scores, and sub-scales scores from three interim assessments. The researchers found all three interim assessments were statistically significant to be a strong predictor of state summative test performance with scale scores being the strongest predictor variable. For example, the results for Interim Assessment 1 analysis revealed that student demographics $F(3,3675) = 203.44$, $p < .001$ and scale scores $F(4, 3675) = 4{,}916.85$, $p < .001$ were statistically significant, whereas sub-scale scores were not found to be statistically significant $F(15, 3660) = 1.38$, $p = .18$.

Overall, scale scores were the strongest predictor explaining approximately 64% of the variance in state summative test performance (Immekus & Atitya, 2016). Additionally, the researchers discuss the importance of construct and alignment of assessment systems to state academic standards to attain the strongest relationship of predictability of student performance

levels. Decisions regarding the predictive use of computerized adaptive assessment scores should be based on total scale scores, not subscales, and closely aligned with state academic standards to be deemed a strong valid predictor of student performance levels on summative high-stakes tests. More importantly, this study suggests computerized adaptive assessments can serve in the capacity of a formative computerized adaptive assessment when aligned to state academic standards to provide data-based evidence for data-based decision-making (Van der Kleij et al., 2015) for predictive purposes related to student performance levels on state summative standardized tests.

A similar study was conducted to determine if a computerized adaptive math assessment, iReady math scores could be used to predict student performance on End-of-Course high-stakes math exams (Thompson, 2018). This study was based on tenth-grade students' achievement scores (N=220) on the iReady math adaptive assessment while controlling for gender, ethnicity, and socioeconomic status (SES). The analysis revealed iReady math scores to be a significant, positive predictor of End-of-Course math exam scores with $\beta = .64$ ($p < .001$) with an $R^2 = .43$. This suggests that 43% of the variance in End-of-Course math exams could be explained by the predictor variable, the iReady math assessment. Additionally, the analysis revealed that the controlling variables did not have a significant moderating influence and suggest that the iReady math adaptive assessment would constitute a fair and unbiased assessment by gender, ethnicity, or SES.

The results of this study suggest that students who do not perform well on the iReady math computerized adaptive assessment due to a lack of skill knowledge and competency do not perform well on End-of-Course criterion exams allowing time for interventions before End-of-Course exams (Thompson, 2018). The findings support the use of technology-based online

adaptive assessment systems for instant accessibility of student data thus providing educators with timely and adequate student data for effective decision making at the classroom, school, and district levels (Choi & McClenen, 2020). Additionally, the findings suggest that the iReady math adaptive assessment is a fair and unbiased assessment, which is a vital component of test validity (Bennett, 2015). Lastly, this study supports the use of formative computerized adaptive assessment systems, such as iReady, as a strong predictor of student performance level on state summative tests (Immekus & Atitya, 2016).

Another study was conducted to examine the relationship of formative online diagnostic assessment system scores as a predictor of student performance levels on state standardized test scores in ELA and math for Grades 4-6 using 2018-2019 archival testing data (Bruce, 2019). This research study was based on mid-year student scores from the iReady online diagnostic assessment system for reading and math and student performance scores on the Missouri Assessment Program (MAP), which are state summative standardized tests for students in Grades 4-6. Student assessment data ranged from fifth-grade reading (N=398) to sixth-grade math (N=416) with total reading scores (N=1,216) and math scores (N=1,223). The analysis revealed that the student scores on the iReady diagnostic assessments in reading and math for Grades 4-6 to be a statistically significant predictor of the student performance test scores on the MAP with an R^2 range from .58 on fourth-grade math to .71 on fifth-grade reading. In other words, this suggests that on the low end 58% of the variance in Grade 4 math MAP scores could be explained by the predictor variable, iReady Grade 4 diagnostic assessment, and on the high end, 71% of the variance in Grade 5 math MAP scores could be explained by the predictor variable, iReady Grade 5 reading diagnostic assessment. Additionally, the analysis revealed that there was a statistically significantly positive, high correlation at or above 0.76 of academic

standards alignment between the iReady diagnostic assessment system and the MAP in Grades 4-6 in reading and math.

This research study suggests that online formative computerized adaptive assessment systems, such as the iReady diagnostic adaptive assessment system for reading and math, can serve as a strong predictor of student performance level on state standardized summative tests when highly aligned to state academic standards (Bruce, 2019; Immekus & Atitya, 2016). Additionally, this study supports that some formative computerized adaptive assessment systems can serve multiple purposes (Immekus & Atitya, 2016). As such, formative computerized adaptive assessment systems can be utilized and function as a diagnostic screening tool and as a formative assessment system to assess student learning and progress toward meeting state academic standards as measured on state summative standardized tests, by pinpointing student learning on a specified learning progression and identifying students at risk of not meeting grade-level proficiency on state summative tests, which allows educators to intervene and adjust instruction before summative testing (Bennett. 2015; Collares & Cecilio-Fernandes, 2019; Shapiro et al., 2015; Thompson, 2018; Van der Kleij et al., 2015). Moreover, this study suggests and supports with predictive validity and valid evidence that a formative computerized adaptive assessment system can utilize student tests scores for predictive purposes to guide decision-makers at the classroom, school, and district levels (Immekus & Atitya, 2016; Perie et al., 2009; Perie et al, 2007) and to assist education leaders and policymakers in their decision-making processes (Van der Kleij et al., 2015) when seeking to establish or maintain an effective comprehensive assessment system that provides a stream of real-time assessment data (Choi & McClenen, 2020; Immekus & Atitya, 2016; Stiggins & DeFour, 2009).

Local Formative Computerized Tests

The use of online computerized testing systems is an increasingly common assessment practice in education due to their proven efficiency and flexibility, which enables districts, schools, and classrooms to manage, measure, predict, maximize student learning and achievement, and report real-time student data through analytics (Choi & McClenen, 2020). On-going data collection and data-driven decision-making are essential to guide educators' decision-making to promote and ensure student learning and achievement. Given the pressure of educators to raise student performance levels on high-stakes accountability tests, many school districts use multiple formative computerized assessment systems. District leaders, school leaders, and teachers are continuously making crucial decisions via data gleaned from these assessments (Van der Kleij et al., 2015).

A new study by The United Nations Educational, Scientific, and Cultural Organization (UNESCO, 2021) predicts “that over 100 million children will fall below proficiency level in reading due to the Covid-19 school closures” (para. 1). Before the pandemic, the number of children lacking basic reading skills was already declining and predicted to fall from 483 million to 460 million. As a result of the pandemic, the number of students lacking basic reading skills jumped to 584 million in 2020, which is an increase of more than 20% and erased previously made gains through education efforts. The report finds that without exceptional efforts and strategies made to mitigate learning loss it could take ten years to recover to pre-pandemic status. The Covid-19 pandemic disruptions in education globally and locally exacerbate the usefulness of gauging student learning and achievement using formative computerized assessments for predictive purposes to meet the needs of students and for educators to optimize student learning across a wide educational spectrum and in compliance with accountability measures.

The use of online formative computerized assessments is of particular interest since the COVID-19 outbreak in 2020; therefore, formative assessments used in this specific East Tennessee school district utilized online formative assessments from multiple sources. They included the iReady Diagnostic Assessments for Tennessee and the Tennessee Mock Interim Assessments for predictive purposes during the 2020-2021 school year. The 2020-2021 school year was amid the reopening of schools due to the Covid-19 pandemic outbreak that closed schools during the 2019-2020 school year.

iReady Diagnostic Assessments for Tennessee

The iReady Diagnostic Assessment program is a computerized adaptive assessment system designed to provide educators with actionable insight into student needs and is based on Tennessee state academic standards (Curriculum Associates, 2021a). Diagnostic assessments offer a complete picture of student performance and growth, eliminating the need for multiple, redundant tests. iReady Diagnostic assessment system provides a suite of data reports that transform data into meaningful and actionable insights for differentiating instruction through pinpointing student strengths and knowledge gaps with targeted instructional tools. iReady programs are commercially owned and published by Curriculum Associates for K-12 educational use.

Curriculum Associates (2021b) programs consist of iReady Assessments and iReady Learning, which include computer-based assessments and online personalized instructional modules in reading and math based on the student performance on the diagnostic assessments. Customers may choose to purchase the entire iReady program or the iReady Assessment tools. With the iReady Assessments program, educators have access to diagnostic assessments, growth monitoring assessments, reports, and other educator assessment resources to provide valid and

reliable information on student learning, mastery, and growth. The iReady Assessments consist of:

> *iReady Diagnostic* for K-12 grades covering Reading and Mathematics in a digital, adaptive form,
>
> *iReady Standards Mastery* for Grades 2-8 covering Reading and Mathematics in a digital, fixed form,
>
> *iReady Assessment of Spanish Reading* for Grades K-6 in a digital, fixed form,
>
> *iReady Oral Reading Fluency Assessments* for Grades 1-6, which are administered offline,
>
> *iReady Early Reading Tasks* for Grades K-3+ that are administered offline, one on one,
>
> *iReady Dyslexia Screener* for Grades K-3 pairing the i-Ready Diagnostic with offline, one-on-one administered assessments,
>
> *Tools for Instruction* for actionable, in-the-moment lesson plans to address gaps identified by the Diagnostic,
>
> Unparalleled service and ongoing educator support from i-Ready Partners. (Curriculum Associates, 2021a, para. 1)

Student personalized learning modules are packaged separately. Additionally, the Curriculum Associates website provides a plethora of information about their products including research studies conducted by and for Curriculum Associates (Curriculum Associates, 2021b).

According to Curriculum Associates (2020), there is a strong correlation between iReady Diagnostic Assessment for Tennessee test scores and TCAP test scores based on 2016-2017 school year data, which was derived from a study by Curriculum Associates in partnership with the Educational Research Institute of America (ERIA). This study shows a correlational average

for Grades 3-8 in reading as .78 and .83 for math. Curriculum Associates (2022) released a more recent correlational linking study between iReady Spring Diagnostic Assessments for Tennessee test scores and the 2020-2021 TCAP test scores for reading and math. This study collected data from approximately 63,000 students from 12 school districts across Tennessee. The school districts were selected specifically to represent the state in terms of the following factors: urbanicity, race/ethnicity, and socioeconomic status (using the National School Lunch Program as a proxy). This study was based on students' test scores for reading in Grades 3-8 (N= 34,000) and math test scores in Grades 3-8 (N= 54,000). The study found a strong correlation between iReady Diagnostic Spring Assessment scores and TCAP test scores administered during the 2020-2021 school year in reading and math. This study shows a correlational average for Grades 3-8 in reading as .81 and .83 in math for the iReady Spring Diagnostic. A correlation of .70 or higher is considered to be a strong correlation between the assessment constructs, which is commonly accepted evidence of validity by the National Center on Intensive Intervention (NCII

The Curriculum Associates (2021b) website lists several linking studies to various states' summative state tests marking high correlations to each respectively. There are critics of Curriculum Associates' correlational studies to state academic standards due to the studies being commissioned by Curriculum Associates calling the objectivity of the studies questionable, such as in Tennessee's linking study by ERIA, due to the linking studies' lack of peer-reviewed journals (Bjorklund-Young & Borkoski, 2016). Additionally, beyond the previously cited iReady assessment research studies (see cited Bruce, 2019; Thompson, 2018), there have been other research studies conducted surrounding the use of iReady's Learning program that showed mixed results (see references Aguilar, 2019; Croteau, 2011; Neligan, 2020; Torres, 2019); however, they are not related to the intent of this research study

Tennessee Mock Interim Assessments

The Tennessee Department Of Education (2020) released free district resources and planning guides for assessing student learning via the Schoolnet platform for Grade 3 and up (Tennessee Department of Education, 2020a). Schoolnet is an instructional improvement system owned and published by Pearson (Pearson, 2015). The state department launched the Schoolnet platform in the fall of 2020, which is a formative assessment platform with online formative assessment tools aligned to TCAP summative tests and the Tennessee Academic Standards (Tennessee Department of Education, 2020a). The district planning guides and formative assessment tools were created to support schools and district-level plans in measuring how students are learning, progressing, performing, and mastering skills toward meeting grade-level expectations by pinpointing and identifying students' learning strengths and weaknesses and providing a pathway to adjust instructional practices to support ongoing learning (Tennessee Department of Education, 2020a; Tennessee Department of Education, 2020b). These are part of the TDOE Innovative Assessment Plan and reopening schools' resources due to COVID-19 school closures in the spring of 2020. "It is critical that educators and schools can identify student progress early and receive actionable data for the upcoming year" (Tennessee Department of Education, 2020a. para. 2).

There are three optional tools available to districts: Start of the Year Checkpoint exams, Online Formative Assessment Platform and TCAP Item Bank, and Full-length "Mock" Interim Assessment (Tennessee Department of Education, 2020a). These resources are provided to Tennessee school districts at no cost. Additionally, each of the assessments may be administered by educators via an online computerized-based testing format or pencil-paper testing format, which increases flexibility for school district use.

The Tennessee Mock Interim Assessments are complete blueprint-aligned assessments that "mirror the current TCAP summative assessments, as well as provide accurate scaled scores and performance bands that will estimate each student's performance" (Tennessee Department of Education, 2020b, p.12). According to TDOE (2020a), the Full-length “Mock” Interim Assessment will:

> Yield much better information for educators than a practice exam, these assessments are scheduled to be available for optional administration to students as a mid-year or spring checkpoint. To support instruction, each item will be released with a full item analysis to assist educators in pinpointing student misconceptions within standards to support efficient re-teaching, review, or remediation. (para. 9)

Additionally, the Mock Interim Assessments are based on the previous TCAP tests, which allows the state department to provide immediate feedback and they are developed identically to TCAP administration providing students with a more authentic practice for future TCAP tests (Tennessee Department of Education, 2020b). The Schoolnet platform dashboard provides educators at various levels in a school system with student data from the Mock Interim Assessments that will enable data-informed decisions regarding relevant data analysis, tasks, and resources to support and promote student learning and achievement (Pearson, 2015). According to the Tennessee Department of Education (2021d), the school districts that used the state department’s free, formative assessment tools showed higher success rates.

Program Evaluation

As federal and state money is allocated and dispensed, there is public and policymaker interest in the return of investment regarding tax dollars. “Policymakers consistently have wanted to know whether federal education programs make a difference in how teachers teach

and how much students learn" (Birman & Porter, 2002, pp. 58-59). Educators can read about their schools and school districts' successes or failures in the newspaper or online. Public reporting of yearly mandated state standardized test scores reflects the school's ability to increase student learning and achievement. This necessitates a priority of focus on program evaluation of computerized adaptive assessment systems used to monitor student learning, achievement, and predictability of student performance level on high-stakes tests.

In education, evaluation and research differ, although there may be overlapping between the two fields to an extent. Education research seeks to add to the existing knowledge base and develop an understanding without judgment, whereas evaluation in education seeks useful information to decision-makers regarding the quality, value, or merit of the worth of the program or product using judgment (Fitzpatrick et al., 2011). Additionally, the results in evaluations are usually limited to the evaluation's purpose, which limits generalizability. On other hand, research offers to generalize findings of the relationships among variables in different settings thus offering more generalizability.

According to Fitzpatrick et al., (2011), mainstream program evaluation is already within many schools today and is considered "part of the organization's work ethic, its culture, and job responsibilities at all levels of the organization … is the process of making evaluation integral part of an organization's everyday operations" (p. 236). Mainstream program evaluation uses guides of internal empowerment in decision-making and action on program planning and improvement as related to the teaching and learning cycle, which is often equated to high-stakes summative state test scores.

There are two types of program evaluation, formative and summative. Scriven (1967, as cited in Fitzpatrick et al., 2011) made the first distinction between the two regarding their

functions. Formative evaluation provides information about the functionality of the program and ways to make improvements. Summative evaluation provides information about the associated outcomes as a result of the program's use to aid in decisions of program continuance.

Fitzpatrick et al. (2011) offer two theories of summative program evaluation approaches, the expertise-oriented approach and the consumer-oriented approach, for program evaluation. The expertise-oriented approach relies heavily on the professional judgment of the program being evaluated by content experts. The expertise-oriented approach includes formal and informal reviews, having standards of measurement used to evaluate the program, using single or multiple content experts, and directly affecting the status of the program. Fitzpatrick et al. (2011) suggest that normally there is a team of content experts offering their expertise to produce a sound evaluation of a program or product. The expertise-oriented approach differs from the consumer-oriented approach. The consumer-oriented approach relies heavily on a quantitative method to evaluate the program. The consumer-oriented approach primarily emphasizes determining the value of the product, the importance of validity, measuring quality, determining important criteria, and establishing standards for criteria. However, there can be overlap in formative and summative evaluation purposes and functions since they can serve as primary and secondary purposes to one another (Van der Kleij et al., 2015); thus the same can be applied to expertise- and consumer-oriented approaches purposes and functions respectively.

For effective interim assessment program evaluations, which can include benchmark and diagnostic assessment systems (Perie et al., 2009; Perie et al., 2007), school district leaders ought to develop a theory of action based on the expectations of the product or program (Herman, 2017). Herman (2017) offers six criteria that can be used to evaluate the quality of assessments: alignment, diagnostic value, fairness, technical quality, utility, and feasibility, which are

important characteristics in making purchasing decisions or assessment selection and school districts should demand evidence of each. Lastly, the product or program evaluation process for assessment systems should be rigorous, systematic, and ethical (Fitzpatrick et al., 2011).

According to the National Research Council of Teachers of English (2014), an effective assessment system should employ multiple assessments, be representative of standardized tests, and ensure tests are valid and reliable for student populations. Perie et al. (2009) suggest educational leaders should follow the advice offered in the *Standards for Educational and Psychology Testing* specifically in Standard 11:1:

> Prior to the adoption and use of a published test, the test user should study and evaluate the materials provided by the test developer. Of particular importance are those that summarize the test's purposes, specify the procedures for test administration, define the intended populations of test-takers, and discuss the score interpretations for which validity or reliability data are available. (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 1999, as cited in Perie et al., 2009, p. 11)

Given the pressure of educators to raise student performance levels on high-stakes accountability tests and the attraction of using formative assessments for predictive purposes by educational institutions, some test publishers began relabeling many of their testing packages as 'formative" (Popham, 2006b). Most people think in terms of the outcomes regarding a product or program evaluation, especially when advertised as researched-based, research-tested, or research-proven; therefore, determining the criteria for product or program evaluation is reliant on the product's intent and the purpose of the evaluation in determining the scope, subject, and methods by the evaluator (Fitzpatrick et al., 2011). When considering research-based programs,

materials, or products as a consumer, educational leaders must seek the purpose and effectiveness of the product or program, while looking past the research-based advertisement, to examine the validity and reliability of the program, material, or product.

Theoretical Framework

Currently, in education, there is a mix of formative assessment practices and approaches implemented to find quick fixes, yet there are limited studies that relate formative assessments to learning theories (Van der Kleij et al., 2015). More importantly, "One of the most neglected aspects of CAT is its alignment with modern learning theories" (Collares & Cecilio-Fernandes, 2019, p. 115). The theoretical foundational basis for this study proposes several learning theories that influence the utility and functionality of formative computerized adaptive assessments for educators' decision-making processes in that formative assessments lie at the intersection of instruction and assessments within the teaching and learning cycle (Van der Kleij et al., 2015). This requires assessment alignment with an understanding of formative computerized adaptive assessments' functionality and utility with theoretical underpinnings for educators' prudent data-driven decision-making processes.

Skinner's Behaviorism Theory is based on observable and measurable facets of behaviors (Zhou & Brown, 2015). Behaviorists view behaviors as conditioning for positive behaviors while deterring unwanted or negative behaviors through stimulus and response mechanisms where every action has a consequence. If the behavior results in good consequences, reinforcements can be used effectively to encourage desired behaviors while withdrawal, punishments, or negative reinforcements can deter unwanted behaviors. Behaviorists view motivation for learning relative to positive and negative reinforcements within the environment and culture, which can be observed and measured. Behaviorists assess the degree of learning using methods that measure

observable behavior such as exam performance using formative computerized assessments that are aligned to state academic standards (Bennett, 2015; Immekus & Atitya, 2016; Lang & Tay, 2020; Shapiro et al., 2015; Zanon et al., 2016).

Piaget's theory of Cognitive Development is based on the belief that learners construct their thoughts and understanding of the world around them through their experiences and that learning is structured in levels. The learner passes through each stage of development incrementally to advance to the next stage of cognitive development (Zhou & Brown, 2015). Constructivist views learning as information processing and representation of knowledge in memory (Van der Kleij et al., 2015) and learning and meaning as an ongoing process built upon prior knowledge of relationships and experiences within the environment in constructs called schema. (Zhou & Brown, 2015). This corresponds to using assessments to identify students' level of achievement and predict student proficiency of performance levels on high-stakes tests utilizing formative computerized adaptive tests since test items are uniquely designed to elicit and match student responses characterized by specified learning progressions based on mandated state academic standards (Bennett, 2015; Choi & McClenen, 2020; Immekus & Atitya, 2016), which represent information processing and store knowledge in students' memory. The use of formative computerized adaptive assessments in this manner not only constitutes an assessment for learning but also an assessment of learning (Collares & Cecilio-Fernandes, 2019).

Bandura's theory of Social Cognitive Theory (SCT) is based on the idea that people learn from observing others through modeling and intimating (Zhou & Brown, 2015). The SCT's emphasis is on the interlocking interplay of how cognitive factors and the environment interact to influence learning and behavior reciprocally where the learner is an active participant through self-regulation and reflection. The SCT is often referred to as a bridge between behaviorist and

cognitive learning theories. Bandura's SCT posits learners can consciously change and develop when the new behavior goes through four steps: attention, retention, reproduction, and motivation are associated with consequential actions positively. In line with the SCT, providing students with the opportunity for assessment practice that mimics high-stakes summative standardized tests based on a stipulated learning trajectory can provide students with successful testing experiences by reducing the cognitive stress anxiety that will promote self-efficiency for future tests, especially for students with tests anxiety (Nagandla et al., 2018). Practice with CATs in a less stressful testing environment minimizes the cognitive load (Collares & Cecilio-Fernandes, 2019), which is ideal and affords all students the exposure and opportunity to be active participants in evaluating their learning to make necessary cognitive and behavior changes based on their test scores as well as providing educators assessment data to assist students' regulation and reflection processes in preparation for future testing (Nagandla et al., 2018; Styron & Styron, 2012; Van der Kleij et al., 2015).

Vygotsky's Sociocultural Theory is based on the relationship between the social world and cognitive development and their interactions in the context of culture as a motivational means for learning to occur (Zhou & Brown, 2015). Vygotsky's Sociocultural Theory for cognitive development suggests that social interactions lead to continuous step-by-step changes in children's development, which can vary from culture to culture (Vygotsky, 1978). Vygotsky (1978) theorized the model of the Zone of Proximal Development, which focuses on the interaction between student learning and performance in a social context. Vygotsky (1978) defines ZPD as: "the distance between the actual developmental level as determined by independent problem solving and the level of potential development as determined through problem-solving under adult guidance or in collaboration with more capable peers (p. 86).

Concerning this study, Vygotsky's theory of ZPD posits that using a formative computerized adaptive assessment method of predicting student performance levels on high-stakes standardized summative tests by matching test difficulty to students' ability (Collares & Cecilio-Fernandes, 2019) can provide educators with an approximation of students' ZPD by using student test scores to improve student performance on high-stakes tests beforehand (Collares & Cecilio-Fernandes, 2019; Van der Kleij et al., 2015). This places high importance on educational leaders to ensure the selection and use of the best formative computerized adaptive assessment system that will produce and provide valid and reliable assessment data regarding their student population (Bruce, 2019; Immekus & Atitya, 2016; Thomas, 2018; Shapiro et al., 2015).

For efficient and effective data-driven decision-making using formative computerized adaptive assessment practices when viewed as an assessment approach, education leaders must encapsulate supportive learning theories that focus on the learning process and learning outcomes simultaneously (Collares & Cecilio-Fernandes, 2019; Van der Kleij et al., 2015). Education leaders must have an underpinning theoretical foundational basis for understanding how various learning theories have different relevance at different stages of learning (Van der Kleij et al., 2015) and the utility of formative assessment data gleaned to fully maximize the interrelated complexity of teaching, learning, and assessments. Thus the integration of learning theories may offer support and the functional utility of formative computerized adaptive assessments as a formative assessment approach to serve as a vehicle for education leaders' data-driven decision-making processes in that the "overarching formative assessment and formative evaluation approach could lead to more valid decisions at different levels in schools … in a thoughtful way can lead to more valid formative decisions" (Van der Kleij et al., 2015, p. 15).

Chapter Summary

The literature review supports significant relationships between federal and state governments' role in public education and student test scores on state-mandated standardized tests (McGuinn, 2015). Although the academic standards and accountability systems have changed in Tennessee, there remains an intensive focus on raising student standardized summative test scores via the state's comprehensive assessment system. While assessments have been a part of the educational system for more than 100 years, educational assessment methods and testing formats have changed over time (Shepard, 2016). Therefore, the need for education leaders to identify the right type of formative computerized assessment system can be daunting as pressure continues to mount through high-stakes testing within the realm of accountability and political ramifications. It is imperative educational decision-makers have the most recent available data from online formative assessment systems currently used to make the best decision for their respective student population regarding functionality, validity, and reliability.

This literature review adds to the existing body of research on formative assessment systems in that formative computerized adaptive assessments can serve as an approach to improve student learning, achievement, and performance level on state exams. When viewed as an approach, formative computerized adaptive assessments can be used to provide educators with real-time data-based evidence of student learning, progress, and achievement in mastering skills and knowledge assessed on state standardized tests to optimize student learning beforehand (Stiggins, 2005; Wiliam, 2014; Van der Kleij et al., 2015). It is vital to assess students' performance and progress throughout the school year and utilize the formative assessment data to tailor instruction reiterating the significant relationships between formative computerized adaptive assessment systems and measuring student learning and progress for predictive

purposes (Bruce, 2019; Immekus & Atitya, 2016; Sutter et al., 2020; Thomas, 2018). For formative computerized adaptive assessment systems to be truly effective at the district and school level, they must be able to provide different kinds of informal assessment data to various policy and decision-makers at different times in various forms for three different levels the classroom, the school, and the district to optimize student learning (Immekus & Atitya, 2016; Stiggins & DeFour. 2009; Van der Kleij et al., 2015).

Additionally, this literature review shows the importance for education leaders to consider the design of formative computerized assessment systems that will adequately fulfill the more important purpose first. Then they can consider if the same assessment can fulfill additional purposes or if multiple assessments are needed for a comprehensive assessment system (Perie et al., 2009). Identifying the goal of the assessment's intended purposes and ensuring that the assessments meet the criteria of intended purposes must be at the forefront as education leaders continue to seek for and use formative computerized tools and systems (Fitzpatrick et al., 2011; Herman, 2017; Perie et al., 2009). Formative computerized assessment systems must be aligned to state academic standards to function as a reliable measurement means of student learning and for student learning and to improve student performance levels towards meeting the goals of state academic standards assessed on state exams annually beforehand (Bruce, 2019; Immekus & Atitya, 2016; Thomas, 2018).

When purchasing or selecting formative computerized assessment systems, such as iReady Diagnostic Assessments or Tennessee Mock Interim Assessments, school district leaders ought to develop a theory of action based on the expectations of the product or program to improve student learning (Herman, 2017). For interim assessments, Herman (2017) offers six criteria that can be used to evaluate the quality of assessments: alignment, diagnostic value,

fairness, technical quality, utility, and feasibility, which are important characteristics in making purchasing decisions or assessment selection and school districts should demand evidence of each. Lastly, the product or program evaluation process for assessment systems should be rigorous, systematic, and ethical (Fitzpatrick et al., 2011).

Although standards-based education reforms will continue to evolve and will result in changes in curriculum, instruction, and assessment, the effectiveness of these efforts will nonetheless rest on the technical quality of the formative computerized adaptive assessment data (Immekus & Atitya, 2016; Van der Kleij et al., 2015). As such, there is a continued need to gather valid and reliable evidence to support the use and interpretation of formative computerized adaptive assessment scores by teachers and administrators to effectively guide classroom, school, and district-level decisions to optimize student learning (Collares & Cecilio-Fernandes, 2019; Immekus & Atitya, 2016; Van der Kleij et al., 2015).

CHAPTER 3

Research Design and Methods

There is evidence that the use of formative assessments guide and inform instruction that leads to gains in student learning and student achievement (Black & Wiliam, 1998a; 1998b). Due to modern accountability systems, educators use formative computerized assessments to measure the effectiveness of the teaching and learning cycle to increase student learning and achievement more intensely. Accountability systems are based largely on state standardized test scores. Accountability trends may explain why there is more formative assessment data in schools. However, the question often remains which type formative computerized assessment produces quality data to assess what students are learning and to what degree students are progressing toward achieving state academic standards' goals and higher performance levels on annual high stakes summative tests.

For online formative assessments and systems to be truly effective at the district level, they must be able to provide different kinds of information to various policy and decision-makers at different times in various forms for three different levels. These levels include the classroom, the school, and the district (Stiggins & DeFour, 2009). It is imperative that education decision-makers have the most recent available data from online formative assessment systems currently used to make better-informed decisions for their respective student population regarding functionality, validity, and reliability. This chapter examines the research design and methods of the research, followed by the research questions and null hypotheses, population and sample, instrumentation, data collection and procedures, and data analysis.

Research Questions and Null Hypotheses

Six major research questions and their subset questions and the null hypotheses were used to guide the collection and analysis of data for this research.

1. Is there a significant relationship between iReady Diagnostic scores and 3rd-grade TNReady ELA test scores?

   $H_0$1: There is no significant relationship between iReady Diagnostic scores and 3rd-grade TNReady ELA test scores.

1a. Is there a significant relationship between Mock Interim scores and 3rd-grade TNReady ELA test scores?

   $H_0$1a: There is no significant relationship between Mock Interim scores and 3rd-grade TNReady ELA test scores.

1b: Is iReady Diagnostic Assessments or Mock Interim Assessments the best predictor of 3rd-grade TCAP ELA test scores?

   $H_0$1b: There is no best predictor of 3rd-grade TCAP ELA test scores by iReady Diagnostic Assessments or Mock Interim Assessments.

2. Is there a significant relationship between iReady Diagnostic scores and 3rd-grade TNReady math test scores?

   $H_0$2: There is no significant relationship between iReady Diagnostic scores and 3rd-grade TNReady math test scores.

2a: Is there a significant relationship between Mock Interim scores and 3rd-grade TNReady math test scores?

   $H_0$2a: Is there a significant relationship between Mock Interim scores and 3rd-grade TNReady math test scores?

2b: Is iReady Diagnostic Assessments or Mock Interim Assessments the best predictor of 3rd-grade TCAP math test scores?

$H_0$2b: There is no best predictor of 3rd-grade TCAP math test scores by iReady Diagnostic Assessments or Mock Interim Assessments.

3. Is there a significant relationship between iReady Diagnostic scores and 4th-grade TNReady ELA test scores?

$H_0$3: There is no significant relationship between iReady Diagnostic scores and 4th-grade TNReady ELA test scores.

3a. Is there a significant relationship between Mock Interim scores and 4th-grade TNReady ELA test scores?

$H_0$3a: There is no significant relationship between Mock Interim scores and 4th-grade TNReady ELA test scores.

3b. Is iReady Diagnostic Assessments or Mock Interim Assessments the best predictor of 4th-grade TCAP ELA test scores?

$H_0$3b: There is no best predictor of 4th-grade TCAP ELA test scores by iReady Diagnostic Assessments or Mock Interim Assessments.

4. Is there a significant relationship between iReady Diagnostic scores and 4th-grade TNReady math test scores?

$H_0$4: There is no significant relationship between iReady Diagnostic scores and 4th-grade TNReady math test scores.

4a. Is there a significant relationship between Mock Interim scores and 4th-grade TNReady math test scores?

$H_0$4a: There is no significant relationship between Mock Interim scores and 4th-grade TNReady math test scores.

4b. Is iReady Diagnostic Assessments or Mock Interim Assessments the best predictor of 4th-grade TCAP math test scores?

$H_0$4b: There is no best predictor of 4th-grade TCAP math test scores by iReady Diagnostic Assessments or Mock Interim Assessments.

5. Is there a significant relationship between iReady Diagnostic scores and 5th-grade TNReady ELA test scores?

$H_0$5: There is no significant relationship between iReady Diagnostic scores and 5th-grade TNReady ELA test scores.

5a. Is there a significant relationship between Mock Interim scores and 5th-grade TNReady ELA test scores?

$H_0$5ba: There is no significant relationship between Mock Interim scores and 5th-grade TNReady ELA test scores.

5b. Is iReady Diagnostic Assessments or Mock Interim Assessments the best predictor of 5th-grade TCAP ELA test scores?

$H_0$5b: There is no best predictor of 5th-grade TCAP ELA test scores by iReady Diagnostic Assessments or Mock Interim Assessments.

6. Is there a significant relationship between iReady Diagnostic scores and 5th-grade TNReady math test scores?

$H_0$6: There is no significant relationship between iReady Diagnostic scores and 5th-grade TNReady math test scores.

6a. Is there a significant relationship between Mock Interim scores and 5th-grade TNReady math test scores?

$H_0$6a: Is there a significant relationship between Mock Interim scores and 5th-grade TNReady math test scores?

6b. Is iReady Diagnostic Assessments or Mock Interim Assessments the best predictor of 5th-grade TCAP math test scores?

$H_0$6b: There is no best predictor of 5th-grade TCAP math test scores by iReady Diagnostic Assessments or Mock Interim Assessments.

Population and Sample

The population of this study was 1,185 students who were in Grades 3 through Grade 5 in 2020-2021 from seven elementary schools in one East Tennessee school district. In Grade 3, there were 402 students. In Grade 4, there were 359 students. In Grade 5, there were 416 students.

The sample for this study was limited to 995 students from seven elementary schools who were in Grades 3-5 in 2020-2021 in the East Tennessee school district. The sample consisted of students who were administered the iReady Diagnostic Assessments, the Tennessee Mock Interim Assessments, and the TCAP tests for reading and math in the 2020-2021 school year. The demographic profile for the participants in the study is displayed in Table 1.

Table 1

Demographic Profile for the Participants in the Study

Grade	(N)	ELA (n)	Math (n)
3	402	330	326
4	367	305	303
5	416	353	360

Instrumentation

The data used in this study were collected using iReady Diagnostic for Tennessee, the Tennessee Mock Interim Assessments, and TCAP TNReady assessments for reading and math. The data sets collected were for Grades 3-5 for 2020-2021 from one East Tennessee school district.

iReady Diagnostic Assessments for Tennessee

Description. iReady Diagnostic Assessment program is commercially owned and published by Curriculum Associates for K-12 educational use. The iReady Diagnostic Assessment program is a computerized-adaptive assessment system that was developed and designed to provide educators, students, and parents with actionable insight into students' grade-level performance and proficiency levels in reading and math based on Tennessee state academic standards (Curriculum Associates, 2021a). iReady Diagnostic assessments are criterion-referenced multiple-choice tests. The i-Ready Diagnostic assessments were designed for students to obtain approximately 50% of the questions correct and 50% incorrect. The test will find a "just right" student grade-level placement based on student responses.

iReady Diagnostic assessment system provides educators with student data reports. These data reports transform student data into meaningful and actionable insights by pinpointing student strengths and knowledge gaps for differentiating instruction (Curriculum Associates, 2021a). The iReady Diagnostic Assessment system served as the school district's universal screener for Response to Intervention (RTI^2). iReady Reading Diagnostic assesses the following content domains, Foundational Skills, Vocabulary, Comprehension of Informational Text, and Comprehension of Literary Text (i-Ready Central, 2022). The iReady Math Diagnostic assesses the Numbers and Operations, Algebra and Algebraic Thinking, Measurement and Data, and Geometry domains. The iReady Diagnostic assessment system provides student scale scores for both reading and math assessments. The reading and math scale scores provide educators with an overall view of student performance levels relative to expectations set by Tennessee Academic Standards.

Administration. Students take the iReady Diagnostic assessments three times a year, fall, winter, and spring. For this research study, only the spring iReady Diagnostic test scores were used. The iReady Spring Diagnostic Assessments were administered to students in Grades 3-5 from April 6 to May 14 with an extension no later than May 21 due to COVID-19 related guidelines and quarantines. Students in this East Tennessee school district had the option of in-person learning or virtual learning due to the reopening of schools during the Covid-19 pandemic. However, all students were required to take the iReady Diagnostic reading and math assessments in person at their home school during the testing window. The iReady Diagnostic assessments were not timed since the assessments are adaptive and differentiated based on individual students' performance.

Scoring. Student scale scores range from 100-800 for all iReady Diagnostic assessments. Student criterion-referenced scale scores are aligned with grade-level placement charts for both reading and math. The grade-level placement charts visually and numerically represent students' performance levels and progress towards growth measures and proficiency levels that are relative to Tennessee Academic Standards. The grade-level placement charts provide educators with data on students' performance levels towards meeting on-grade level expectations on state standardized tests. The iReady Spring Diagnostic Grade-Level placements for Grades 3-5 reading and math are detailed in Table 2. Each grade-level placement has a defined on-grade level range. The iReady Spring Diagnostic On-Grade level ranges for Grades 3-5 reading and math are displayed in Table 3.

Table 2

iReady Spring Diagnostic Grade-Level Placements for Grades 3-5

	ELA			Math		
Grade	Below Level	On Level	Above Level	Below Level	On Level	Above Level
3	100-511	511-602	603-800	100-448	449-516	517-800
4	100-556	557-629	630-800	100-464	465-526	527-800
5	100-580	581-640	641-640	100-479	480-540	541-800

Table 3

iReady Spring Diagnostic On-Grade Level Ranges for Grades 3-5

Grade	ELA			Math		
	Early	Mid	Late	Early	Mid	Late
3	511-544	545-560	561-602	449-463	464-506	507-516
4	557-578	579-602	603-629	465-481	482-516	517-526
5	581-608	608-629	630-640	480-497	498-526	527-540

Reliability and Validity. According to Curriculum Associates (2020), there is a strong correlation between iReady Diagnostic Assessment for Tennessee test scores and TCAP test scores based on 2016-2017 school year data, which was derived from a study by Curriculum Associates in partnership with the Educational Research Institute of America (ERIA). The study shows a correlational average for Grades 3-8 in reading as .78 and .83 for math. Curriculum Associates (2022) released a more recent correlational linking study between iReady Spring Diagnostic Assessments for Tennessee test scores and the 2020-2021 TCAP test scores for reading and math. This study collected data from approximately 63,000 students from 12 school districts across Tennessee. The school districts were selected specifically to represent the state in terms of the following factors: urbanicity, race/ethnicity, and socioeconomic status (using the National School Lunch Program as a proxy). This study was based on students' test scores for reading in Grades 3-8 (N= 34,000) and math test scores in Grades 3-8 (N= 54,000). The study found a strong correlation between iReady Diagnostic Spring Assessment scores and TCAP test scores administered during the 2020-2021 school year in reading and math. The study showed a

correlational average for Grades 3-8 in reading as .81 and .83 in math for the iReady Spring Diagnostic. A correlation of .70 or higher is considered to be a strong correlation between the assessment constructs, which is commonly accepted evidence of validity by the National Center on Intensive Intervention (NCII) (Curriculum Associates, 2020). According to Curriculum Associates (2022), this large-scale study provides additional support that i-Ready is a well-researched program that meets the criteria for "evidence-based" that is outlined by the Every Student Succeeds Act (ESSA).

Tennessee Mock Interim Assessments

Description. The Mock Interim Assessments were a part of the Tennessee Department of Education (TDOE) Innovative Assessment Plan and reopening schools' resources due to COVID-19 school closures in the spring of 2020. These formative assessments were created to support schools and district-level plans in measuring how students are learning, progressing, performing, and mastering skills toward meeting grade-level expectations (Tennessee Department of Education, 2020a; Tennessee Department of Education, 2020b). The TDOE (2020a) states, "It is critical that educators and schools can identify student progress early and receive actionable data for the upcoming year" (para. 2).

The Tennessee Mock Interim Assessments were released by the TDOE in the fall of 2020 (Tennessee Department Of Education, 2020) via the Schoolnet platform for Grade 3 and up. The Mock Interim Assessments are online formative assessments aligned to TCAP summative tests and the Tennessee Academic Standards (Tennessee Department of Education, 2020a). The Mock Interim Assessments were designed to pinpoint and identify students' learning strengths and weaknesses. The Schoolnet platform provides educators with student data reports. This provides

a pathway for educators to adjust instructional practices and support ongoing learning. According to TDOE (2020a), the Full-length "Mock" Interim Assessment will:

> Yield much better information for educators than a practice exam, these assessments are scheduled to be available for optional administration to students as a mid-year or spring checkpoint. To support instruction, each item will be released with a full item analysis to assist educators in pinpointing student misconceptions within standards to support efficient re-teaching, review, or remediation. (para. 9)

This enables educators to make data-informed decisions based on relevant student data.

Administration. The Mock Interim Assessments for ELA and math were given during January 2021. Students in this East Tennessee school district had the option of in-person learning or virtual learning due to the reopening of schools during the Covid-19 pandemic. However, all students were required to take the online Mock Interim ELA and math assessments in person at their home school during the testing window. The Mock Interim Assessments were not timed.

Scoring. On the Mock Interim ELA and math test score reports, the raw score ranges were provided and correlated to match the following TCAP TNReady achievement performance levels via a color code: Orange (Level 1- Below); Yellow (Level 2- Approaching); Green (Level 3- On Track); and, Blue (Level 4- Mastered). Cut points were utilized to assign students to these different achievement levels. The cut scores for the 2020-2021 Mock Interim ELA and math assessments for Grades 3-5 are detailed in the following tables: Table 4 shows Grade 3 cut scores; Table 5 shows Grade 4 cut scores; and, Table 6 shows Grade 5 cut scores.

Table 4

Mock Interim Cut Scores for Grade 3

ELA			Math		
Number of Questions	Raw Score Ranges	Performance Levels	Number of Questions	Raw Score Ranges	Performance Levels
38	0-36%	Orange (Level 1- Below)	49	0-35%	Orange (Level 1- Below)
	37-61%	Yellow (Level 2-Approaching)		36-60%	Yellow (Level 2-Approaching)
	62-80%	Green (Level 3-On Track)		61-88%	Green (Level 3-On Track)
	81-100%	Blue (Level 4-Mastered)		89-100%	Blue (Level 4-Mastered)

Table 5

Mock Interim Cut Scores for Grade 4

ELA			Math		
Number of Questions	Raw Score Ranges	Performance Levels	Number of Questions	Raw Scores Ranges	Performance Levels
36	0-38%	Orange (Level 1-Below)	49	0-43%	Orange (Level 1-Below)
	39-63%	Yellow (Level 2-Approaching)		44-68%	Yellow (Level 2-Approaching)
	64-80%	Green (Level 3-On Track)		69-84%	Green (Level 3-On Track)
	81-100%	Blue (Level 4-Mastered)		79-100%	Blue (Level 4-Mastered)

Table 6

Mock Interim Cut Scores for Grade 5

ELA			Math		
Number of Questions	Raw Score Ranges	Performance Levels	Number of Questions	Raw Score Ranges	Performance Levels
36	0-36%	Orange (Level 1- Below)	48	0-38%	Orange (Level 1- Below)
	37-57%	Yellow (Level 2-Approaching)		39-63%	Yellow (Level 2-Approaching)
	58-78%	Green (Level 3-On Track)		64-84%	Green (Level 3-On Track)
	79-100%	Blue (Level 4-Mastered)		85-100%	Blue (Level 4-Mastered)

Reliability and Validity. The Tennessee Mock Interim Assessments are complete blueprint-aligned assessments that "mirror the current TCAP summative assessments, … and provide accurate scaled scores and performance bands that will estimate each student's performance" (Tennessee Department of Education, 2020b, p.12). The Mock Interim Assessments were developed from TCAP tests (Tennessee Department of Education, 2020b). The Mock Interim Assessments are deemed statistically valid since the test questions were derived from the previous TCAP tests. According to the Tennessee Department of Education (2021d), the school districts that used the state department's free, formative assessment tools showed higher success rates.

TCAP (Tennessee Comprehensive Assessment Program) TNReady Assessments

Description. TCAP TNReady ELA and math tests were criterion-referenced, standards-based assessments given to all students in Grades 3-8 based on Tennessee Academic Standards for ELA, math, science, and social studies (Tennessee Department of Education, n.d.e). These state-wide timed standardized assessments meet the ESSA requirements of annual assessments in

ELA, math, science, and social studies. The ELA TNReady tests are subdivided into four subparts that utilize literary and informational text and require students to demonstrate the ability to read closely, analyze text, answer text-dependent questions, provide a written response to a prompt, and demonstrate command of the English language. The math TNReady tests are subdivided into three subparts that require students to demonstrate a deep conceptual understanding of mathematics, number sense, fluency, problem-solving, and an understanding of the grade-level horizontal coherence embedded within the standards. TCAP test items include multiple-choice, multiple select, writing prompts, editing tasks, and evidence-based selected responses. TCAP TNReady tests were created to measure critical thinking and higher expectations for students in Tennessee.

Administration. TCAP TNReady tests were administered to students between April 16 and May 6 of 2021. Additionally, if a student was quarantined after completing a sub-part of the TCAP TNReady test, then make-up tests could be extended to June 10. In the 2020-2021 school year, students in this East Tennessee school district had the option of in-person learning or virtual learning due to the reopening of schools during the Covid-19 pandemic. However, all students were required to take TCAP TNReady tests in person at their home school during the testing window. The total ELA TCAP TNReady tests for all four subparts ranged from 180 minutes to 200 minutes while the total TCAP TNReady math tests for all three subparts were 118 minutes for Grades 3-5.

Scoring. On the TCAP TNReady ELA and math test reports, raw scores and scale scores were provided and correlated to the following achievement performance levels: Level 1- Below (below grade level); Level 2- Approaching (approaching grade level); Level 3- On Track (on grade level); and, Level 4- Mastered (mastered grade level) (Tennessee Department of

Education, n.d.f). The cut scores for the 2020-2021 TCAP TNReady ELA and math assessments for Grades 3-5 are detailed in the following tables: Table 7 shows Grade 3 cut scores; Table 8 shows Grade 4 cut scores; and, Table 9 shows Grade 5 cut scores.

Table 7

TCAP Cut Scores for Grade 3

ELA			Math		
Raw Scores	Scale Scores	Performance Levels	Raw Scores	Scale Scores	Performance Levels
0-19	200-300	1-Below	0-16	200-301	1-Below
20-30	324-358	2-Approaching	17-28	305-340	2-Approaching
31-38	361-387	3-On Track	29-38	343-370	3-On Track
39-52	391-450	4-Mastered	39-50	373-450	4-Mastered

Table 8

TCAP Cut Scores for Grade 4

ELA			Math		
Raw Scores	Scale Scores	Performance Levels	Raw Scores	Scale Scores	Performance Levels
0-15	200-294	1-Below	0-18	200-293	1-Below
16-30	299-342	2-Approaching	19-29	297-328	2-Approaching
31-43	345-377	3-On Track	30-43	331-372	3-On Track
44-51	380-450	4-Mastered	44-52	376-450	4-Mastered

Table 9

TCAP Cut Scores for Grade 5

ELA			Math		
Raw Scores	Scale Scores	Performance Levels	Raw Scores	Scale Scores	Performance Levels
0-21	200-294	1-Below	0-18	200-298	1- Below
22-33	297-332	2-Approaching	19-29	302-338	2-Approaching
34-44	335-369	3-On Track	30-40	341-371	3-On Track
45-54	372-450	4- Mastered	41-52	375-450	4-Mastered

Reliability and Validity. The questions for the TNReady assessments consist of a three-step collaborative evaluation process (Tennessee Department of Education, n.d.b). The first step consists of a collaboration between the state department, Tennessee teachers, and testing vendors to create test questions. The state department sets the expectations for test questions in alignment with the Tennessee Academic Standards, whereas teachers and test vendors develop test questions. The second step consists of a collaboration between state content teams and vendors to ensure alignment and standard expectations. Tennessee teachers' committees are to ensure content alignment and that no bias issue associated with specific student groups or sensitive topics exists, and vendors are to make necessary updates based on the teachers' feedback. The final step consists of the state department and vendors putting the tests together. Each test question is field-tested and reviewed for statistical validity. After test questions are deemed statistically valid, the vendor adds the questions to an operational test followed by a review by the state department for accuracy. Finally, the final TNReady assessment is administered in Tennessee schools.

Data Collection and Procedures

Before conducting this research study, the researcher obtained permission from the Institutional Review Board (IRB) at Milligan University (see Appendix A) based on the research proposal and permission from the school system's Director of Schools (see Appendix B). After permission was granted to begin this study, the researcher met with the school district's Data Supervisor to obtain the following data sets iReady Diagnostic, Mock Interim Assessments, and TCAP scores for ELA and math for students who were in Grades 3 through Grade 5 for the 2020-2021 school year. Data from the school district were collected and organized into a spreadsheet with both student and district identifying information removed and substituted with numerical codes for tracking purposes.

Data Analysis

All data were analyzed using the Statistical Package for Social Sciences (SPSS v. 28). Analyses were conducted for each of the six major research questions and their subset research questions.

1. Research question 1: Pearson correlation coefficient was calculated to determine the relationship between iReady Diagnostic scores and 3rd-grade TNReady ELA test scores.

   Research question 1a: Pearson correlation coefficient was calculated to determine the relationship between Mock Interim scores and 3rd-grade TNReady ELA test scores.

   Research question 1b: Multiple linear regression was calculated to predict performance on 3rd-grade TCAP ELA scores based on iReady Diagnostic Assessments or Mock Interim Assessments scores.

2. Research question 2: Pearson correlation coefficient was calculated to determine the relationship between iReady Diagnostic scores and 3rd-grade TNReady math test scores.

   Research question 2a: Pearson correlation coefficient was calculated to determine the relationship between Mock Interim scores and 3rd-grade TNReady math test scores.

   Research question 2b: Multiple linear regression was calculated to predict performance on 3rd-grade TCAP math scores based on iReady Diagnostic Assessments or Mock Interim Assessments scores.

3. Research question 3: Pearson correlation coefficient was calculated to determine the relationship between iReady Diagnostic scores and 4th-grade TNReady ELA test scores.

   Research question 3a: Pearson correlation coefficient was calculated to determine the relationship between Mock Interim scores and 4th-grade TNReady ELA test scores.

   Research question 3b: Multiple linear regression was calculated to predict performance on 4th-grade TCAP ELA scores based on iReady Diagnostic Assessments or Mock Interim Assessments scores.

4. Research question 4: Pearson correlation coefficient was calculated to determine the relationship between iReady Diagnostic scores and 4th-grade TNReady math test scores.

   Research question 4a: Pearson correlation coefficient was calculated to determine the relationship between Mock Interim scores and 4th-grade TNReady math test scores.

Research question 4b: Multiple linear regression was calculated to predict performance on 4th-grade TCAP math scores based on iReady Diagnostic Assessments or Mock Interim Assessments scores.

5. Research question 5: Pearson correlation coefficient was calculated to determine the relationship between iReady Diagnostic scores and 5th-grade TNReady ELA test scores.

Research question 5a: Pearson correlation coefficient was calculated to determine the relationship between Mock Interim scores and 5th-grade TNReady ELA test scores.

Research question 5b: Multiple linear regression was calculated to predict performance on 5th-grade TCAP ELA scores based on iReady Diagnostic Assessments or Mock Interim Assessments scores.

6. Research question 6: Pearson correlation coefficient was calculated to determine the relationship between iReady Diagnostic scores and 5th-grade TNReady math test scores.

Research question 6a: Pearson correlation coefficient was calculated to determine the relationship between Mock Interim scores and 5th-grade TNReady math test scores.

Research question 6b: Multiple linear regression was calculated to predict performance on 5th-grade TCAP math scores based on iReady Diagnostic Assessments or Mock Interim Assessments scores.

Data analyses were conducted using IBM Statistical Package for Social Sciences (SPSS v. 28). The independent variables were iReady Diagnostic and Mock Interim Assessments, and the dependent variable was student performance levels on the TCAP test scores for students who were in Grades 3 through Grade 5 in ELA and math for the 2020-2021 school year. All data were

analyzed at the .05 significance level. The .05 significance level is the standard of probability that is most commonly used to determine if results are statistically significant (Muijus, 2011). The analysis results for each question are included in chapter four.

Chapter Summary

This chapter contained the research methodology used in this quantitative research study. A brief introduction was followed by the research questions and null hypotheses, and the population and sample were presented. Additionally, the instrumentations used in the research study and the process for data collection and data analysis were described.

CHAPTER 4

Data Analysis and Findings

The purpose of this quantitative study was to examine the relationship between the scores of 3rd–5th-grade students on iReady Spring Diagnostic for Tennessee (iReady Diagnostic) and the Tennessee Mock Interim Assessments (Mock Interim Assessment) and TCAP TNReady summative test scores for ELA and math. A related purpose of this study was to determine the predictive validity of iReady Diagnostic and Mock Interim Assessments scores for ELA and math on Grades 3-5 TCAP TNReady ELA and math performance. Participants in this study included students in Grades 3-5 in 2020-2021 from seven elementary schools in one East Tennessee school district. The students were administered the iReady Spring Diagnostic Assessments, the Tennessee Mock Interim Assessments, and TCAP TNReady tests for reading and math in Grades 3-5 in the 2020-2021 school year. Students who had incomplete data were excluded from the study. This chapter details the findings resulting from the data analysis of the six major research questions and their subset research questions used in this study.

Demographic Data

The population of this study was 1,185 students who were in Grades 3 through Grade 5 in 2020-2021 from seven elementary schools in the East Tennessee school district. In Grade 3, there were 402 students. In Grade 4, there were 359 students. In Grade 5, there were 416 students.

The sample for this study was limited to 995 students from seven elementary schools who were in Grades 3-5 in 2020-2021 in the East Tennessee school district. The sample consisted of students who were administered the iReady Spring Diagnostic Assessments, the Tennessee Mock Interim Assessments, and the TCAP tests for ELA and math in the 2020-2021 school year. Each

student's score on the iReady Spring Diagnostic Assessment and Tennessee Mock Interim Assessment were paired with their score on the TCAP TNReady assessment for ELA and math. Students who had incomplete data were excluded from the study.

Findings

Research Question 1

Research Question 1: Is there a significant relationship between iReady Diagnostic scores and 3rd-grade TNReady ELA test scores?

$H_0$1: There is no significant relationship between iReady Diagnostic scores and 3rd-grade TNReady ELA test scores.

A Pearson correlation coefficient was calculated to determine the relationship between iReady Diagnostic scores and 3rd-grade TNReady ELA test scores. A significant, positive, strong correlation was found [r (330 df) = .76, p = <.05]. To determine the variance explained by iReady Diagnostic scores on 3rd-grade TNReady ELA test scores, a coefficient of determination was computed. The results (r^2= .58) suggest that 58% of the variance in 3rd-grade TNReady ELA tests scores could be explained by the iReady Diagnostic scores. The higher the iReady Diagnostic scores, the higher the 3rd-grade TNReady ELA test scores. As a result of the analysis, the null hypothesis was rejected. The results are displayed in Table 10.

Table 10

Correlation Coefficients of iReady Diagnostic and TNReady ELA Scores for 3rd-Grade

Category	*M*	*r*	*p*	r^2
iReady Diagnostic Scores	518.07	.76	.001	.58
TNReady ELA Test Scores	336.26			

Note. $p < .05$

Research Question 1a: Is there a significant relationship between Mock Interim scores and 3rd-grade TNReady ELA test scores?

$H_0$1a: There is no significant relationship between Mock Interim scores and 3rd-grade TNReady ELA test scores.

A Pearson correlation coefficient was calculated to determine the relationship between Mock Interim scores and 3rd-grade TNReady ELA test scores. A significant, positive, moderate correlation was found [r (330 df) = .66, p = <.05]. To determine the variance explained by Mock Interim scores on 3rd-grade TNReady ELA test scores, a coefficient of determination was computed. The results (r^2= .44) suggest that 44% of the variance in 3rd-grade TNReady ELA tests scores could be explained by the Mock Interim scores. The higher the Mock Interim scores, the higher the 3rd-grade TNReady ELA tests scores. As a result of the analysis, the null hypothesis was rejected. The results are displayed in Table 11.

Table 11

Correlation Coefficients of Mock Interim and TNReady ELA Scores for 3rd-Grade

Category	*M*	*r*	*p*	r^2
Mock Interim Scores	38.67	.66	.001	.44
TNReady ELA Test Scores	336.26			

Note. $p < .05$

Research Question 1b: Is iReady Diagnostic Assessments or Mock Interim Assessments the best predictor of 3rd-grade TCAP ELA test scores?

$H_0$1b: There is no best predictor of 3rd-grade TCAP ELA test scores by iReady Diagnostic Assessments or Mock Interim Assessments.

Multiple linear regression was calculated to predict 3rd-grade TCAP ELA test scores based on iReady Diagnostic and Mock Interim Assessment scores. A significant regression equation was found [$F(2, 327) = 272.454$, $p = .001$] with an R^2 of .625, this suggests that 62.5% of the variance in 3rd-grade TCAP ELA test scores could be explained by the predictor variables. Results also suggest that 37.5% of the variance in 3rd-grade TCAP ELA test scores could be explained by other variables other than the predictor variables. To determine which predictor variables were significant predictors of 3rd-grade TCAP ELA test scores, *Beta* scores were examined. The iReady Diagnostic had a *Beta* score of .577, $p = .001$, and the Mock Interim had a *Beta* score of .281, $p = .001$. Both the iReady Diagnostic and Mock Interim Assessments were significant predictors of 3rd-grade TNReady ELA test scores. However, the iReady Diagnostic

was a stronger predictor than the Mock Interim Assessment of the 3rd-grade TNReady ELA test scores. The null hypothesis was rejected. The results are displayed in Table 12.

Table 12

Coefficients for Each Predictor Variable and the Dependent Variable for 3rd-Grade ELA

Variable	*B*	*Beta*	*t*	*p*
iReady Diagnostic Assessment	.361	.577	12.799	.001*
Mock Interim Assessment	.544	.281	6.229	.001*

Note. *indicates significance at $p < .05$

Research Question 2

Research Question 2: Is there a significant relationship between iReady Diagnostic scores and 3rd-grade TNReady math test scores?

$H_0$2: There is no significant relationship between iReady Diagnostic scores and 3rd-grade TNReady math test scores.

A Pearson correlation coefficient was calculated to determine the relationship between iReady Diagnostic scores and 3rd-grade TNReady math test scores. A significant, positive, strong correlation was found [r (326 df) = .81, p = <.05]. To determine the variance explained by iReady Diagnostic scores on 3rd-grade TNReady math test scores, a coefficient of determination was computed. The results (r^2= .66) suggest that 66% of the variance in 3rd-grade TNReady math tests scores could be explained by the iReady Diagnostic scores. The higher the iReady Diagnostic scores, the higher the 3rd-grade TNReady ELA test scores. The null hypothesis was rejected. The results are displayed in Table 13.

Table 13

Correlation Coefficients of iReady Diagnostic and TNReady Math Scores for 3rd-Grade

Category	*M*	*r*	*p*	r^2
iReady Diagnostic Scores	445.27	.81	.001	.66
TNReady Math Test Scores	320.11			

Note. $p < .05$

Research Question 2a: Is there a significant relationship between Mock Interim scores and 3rd-grade TNReady math test scores?

$H_0$2a Is there a significant relationship between Mock Interim scores and 3rd-grade TNReady math test scores?

A Pearson correlation coefficient was calculated to determine the relationship between Mock Interim scores and 3rd-grade TNReady math test scores. A significant, positive, strong correlation was found [r (326 df) = .80, p = <.05]. To determine the variance explained by Mock Interim scores on 3rd-grade TNReady math test scores, a coefficient of determination was computed. The results (r^2= .64) suggest that 64% of the variance in 3rd-grade TNReady math tests scores could be explained by the Mock Interim scores. The higher the Mock Interim scores, the higher the 3rd-grade TNReady math tests scores. As a result of the analysis, the null hypothesis was rejected. The results are displayed in Table 14.

Table 14

Correlation Coefficients of Mock Interim and TNReady Math Scores for 3rd-Grade

Category	*M*	*r*	*p*	r^2
Mock Interim Scores	41.45	.80	.001	.64
TNReady Math Test Scores	320.11			

Note. $p < .05$

Research Question 2b. Is iReady Diagnostic Assessments or Mock Interim Assessments the best predictor of 3rd-grade TCAP math test scores?

$H_0$2b: There is no best predictor of 3rd-grade TCAP math test scores by iReady Diagnostic Assessments or Mock Interim Assessments.

Multiple linear regression was calculated to predict 3rd-grade TCAP math test scores based on iReady Diagnostic and Mock Interim Assessment math scores. A significant regression equation was found [$F(2, 323) = 460.989$, $p = .001$] with an R^2 of .741, this suggests that 74.1% of the variance in 3rd-grade TCAP math test scores could be explained by the predictor variables. Results also suggest that 25.9% of the variance in 3rd-grade TCAP math test scores could be explained by other variables other than the predictor variables. To determine which predictor variables were significant predictors of 3rd-grade TCAP math test scores, *Beta* scores were examined. The iReady Diagnostic had a *Beta* score of .482, $p = .001$, and the Mock Interim had a *Beta* score of .437, $p = .001$. Both the iReady Diagnostic and Mock Interim Assessments were significant predictors of 3rd-grade TNReady math test scores. However, the iReady Diagnostic was a stronger predictor than the Mock Interim Assessment of 3rd-grade TNReady

math test scores. As a result of the analysis, the null hypothesis was rejected. The results are displayed in Table 15.

Table 15

Coefficients for Each Predictor Variable and the Dependent Variable for 3rd-Grade Math

Variable	*B*	*Beta*	*t*	*p*
iReady Diagnostic Assessment	.644	.482	11.179	.001*
Mock Interim Assessment	.895	.437	10.142	.001*

Note. *indicates significance at p <.05

Research Question 3

Research Question 3: Is there a significant relationship between iReady Diagnostic scores and 4th-grade TNReady ELA test scores?

$H_0$3: There is no significant relationship between iReady Diagnostic scores and 4th-grade TNReady ELA test scores.

A Pearson correlation coefficient was calculated to determine the relationship between iReady Diagnostic scores and 4th-grade TNReady ELA test scores. A significant, positive, strong correlation was found [r (305 df) = .75, p = <.05]. To determine the variance explained by iReady Diagnostic scores on 4th-grade TNReady ELA test scores, a coefficient of determination was computed. The results (r^2= .56) suggest that 56% of the variance in 4th-grade TNReady ELA tests scores could be explained by the iReady Diagnostic scores. The higher the iReady Diagnostic scores, the higher the 4th-grade TNReady ELA tests scores. As a result of the analysis, the null hypothesis was rejected. The results are displayed in Table 16.

Table 16

Correlation Coefficients of iReady Diagnostic and TNReady ELA Scores for 4th-Grade

Category	*M*	*r*	*p*	r^2
iReady Diagnostic Scores	538.54	.75	.001	.56
TNReady ELA Test Scores	323.82			

Note. $p < .05$

Research Question 3a: Is there a significant relationship between Mock Interim scores and 4th-grade TNReady ELA test scores?

$H_0$3a: There is no significant relationship between Mock Interim scores and 4th-grade TNReady ELA test scores.

A Pearson correlation coefficient was calculated to determine the relationship between Mock Interim scores and 4th-grade TNReady ELA test scores. A significant, positive, moderate correlation was found [r (305 df) = .69, p = <.05]. To determine the variance explained by Mock Interim scores on 4th-grade TNReady ELA test scores, a coefficient of determination was computed. The results (r^2= .48) suggest that 48% of the variance in 4th-grade TNReady ELA tests scores could be explained by the Mock Interim scores. The higher the Mock Interim scores, the higher the 4th-grade TNReady ELA tests scores. As a result of the analysis, the null hypothesis was rejected. The results are displayed in Table 17.

Table 17

Correlation Coefficients of Mock Interim and TNReady ELA Scores for 4th-Grade

Category	*M*	*r*	*p*	r^2
Mock Interim Scores	45.10	.69	.001	.48
TNReady ELA Test Scores	323.82			

Note. $p < .05$

Research Question 3b: Is iReady Diagnostic Assessments or Mock Interim Assessments the best predictor of 4th-grade TCAP ELA test scores?

$H_0$3b: There is no best predictor of 4th-grade TCAP ELA test scores by iReady Diagnostic Assessments or Mock Interim Assessments.

Multiple linear regression was calculated to predict 4th-grade TCAP ELA test scores based on iReady Diagnostic and Mock Interim Assessment scores. A significant regression equation was found [$F(2, 302) = 240.552$, $p = .001$] with an R^2 of .614, this suggests that 61.4% of the variance in 4th-grade TCAP ELA test scores could be explained by the predictor variables. Results also suggest that 38.6% of the variance in 4th-grade TCAP ELA test scores could be explained by other variables other than the predictor variables. To determine which predictor variables were significant predictors of 4th-grade TCAP ELA test scores, *Beta* scores were examined. The iReady Diagnostic had a *Beta* score of .537, $p = .001$, and the Mock Interim had a *Beta* score of .307, $p = .001$. Both the iReady Diagnostic and Mock Interim Assessments were significant predictors of 4th-grade TNReady ELA test scores. However, iReady Diagnostic was a stronger predictor than the Mock Interim Assessment of 4th-grade TNReady ELA test scores.

As a result of the analysis, the null hypothesis was rejected. The results are displayed in Table 18.

Table 18

Coefficients for Each Predictor Variable and the Dependent Variable for 4th-Grade ELA

Variable	*B*	*Beta*	*t*	*p*
iReady Diagnostic Assessment	.230	.537	10.669	.001*
Mock Interim Assessment	.419	.307	6.100	.001*

Note. *indicates significance at p <.05

Research Question 4

Research Question 4: Is there a significant relationship between iReady Diagnostic scores and 4th-grade TNReady math test scores?

$H_0$4: There is no significant relationship between iReady Diagnostic scores and 4th-grade TNReady math test scores.

A Pearson correlation coefficient was calculated to determine the relationship between iReady Diagnostic scores and 4th-grade TNReady math test scores. A significant, positive, moderate correlation was found [r (302 df) = .66, p = <.05]. To determine the variance explained by iReady Diagnostic scores on 4th-grade TNReady math test scores, a coefficient of determination was computed. The results (r^2= .44) suggest that 44% of the variance in 4th-grade TNReady math tests scores could be explained by the iReady Diagnostic scores. The higher the iReady Diagnostic scores, the higher the 4th-grade TNReady ELA tests scores. As a result of the analysis, the null hypothesis was rejected. The results are displayed in Table 19.

Table 19

Correlation Coefficients of iReady Diagnostic and TNReady Math Scores for 4th-Grade

Category	*M*	*r*	*p*	r^2
iReady Diagnostic Scores	459.37	.66	.001	.44
TNReady Math Test Scores	306.98			

Note. $p < .05$

Research Question 4a: Is there a significant relationship between Mock Interim scores and 4th-grade TNReady math test scores?

$H_0$4a: There is no significant relationship between Mock Interim scores and 4th-grade TNReady math test scores.

A Pearson correlation coefficient was calculated to determine the relationship between Mock Interim scores and 4th-grade TNReady math test scores. A significant, positive, strong correlation was found [r (302 df) = .73, p = <.05]. To determine the variance explained by Mock Interim scores on 4th-grade TNReady math test scores, a coefficient of determination was computed. The results (r^2= .53) suggest that 53% of the variance in 4th-grade TNReady math tests scores could be explained by the Mock Interim scores. The higher the Mock Interim scores, the higher the 4th-grade TNReady math tests scores. As a result of the analysis, the null hypothesis was rejected. The results are displayed in Table 20.

Table 20

Correlation Coefficients of Mock Interim and TNReady Math Scores for 4th-Grade

Category	*M*	*r*	*p*	r^2
Mock Interim Scores	37.20	.73	.001	.53
TNReady Math Test Scores	306.98			

Note. $p < .05$

Research Question 4b: Is iReady Diagnostic Assessments or Mock Interim Assessments the best predictor of 4th-grade TCAP math test scores?

$H_0$4b: There is no best predictor of 4th-grade TCAP math test scores by iReady Diagnostic Assessments or Mock Interim Assessments.

Multiple linear regression was calculated to predict 4th-grade TCAP math test scores based on iReady Diagnostic and Mock Interim Assessment math scores. A significant regression equation was found [$F(2, 299) = 227.106$, $p = .001$] with an R^2 of .603, this suggests that 60.3% of the variance in 4th-grade TCAP math test scores could be explained by the predictor variables. Results also suggest that 39.7% of the variance in 4th-grade TCAP math test scores could be explained by other variables other than the predictor variables. To determine which predictor variables were significant predictors of 4th-grade TCAP math test scores, *Beta* scores were examined. The Mock Interim had a *Beta* score of .516, $p = .001$, and the iReady Diagnostic had a *Beta* score of .348, $p = .001$. Both the iReady Diagnostic and Mock Interim Assessments were significant predictors of 4th-grade TNReady math test scores. However, the Mock Interim Assessment was a stronger predictor than the iReady Diagnostic of 4th-grade TNReady math test

scores. As a result of the analysis, the null hypothesis was rejected. The results are displayed in Table 21.

Table 21

Coefficients for Each Predictor Variable and the Dependent Variable for 4th-Grade Math

Variable	*B*	*Beta*	*t*	*p*
iReady Diagnostic Assessment	.422	.348	7.647	.001*
Mock Interim Assessment	1.231	.516	11.330	.001*

Note. *indicates significance at p <.05

Research Question 5

Research Question 5: Is there a significant relationship between iReady Diagnostic scores and 5th-grade TNReady ELA test scores?

$H_0$5: There is no significant relationship between iReady Diagnostic scores and 5th-grade TNReady ELA test scores.

A Pearson correlation coefficient was calculated to determine the relationship between iReady Diagnostic scores and 5th-grade TNReady ELA test scores. A significant, positive, strong correlation was found [r (353 df) = .78, p = <.05]. To determine the variance explained by iReady Diagnostic scores on 5th-grade TNReady ELA test scores, a coefficient of determination was computed. The results (r^2= .61) suggest that 61% of the variance in 5th-grade TNReady ELA tests scores could be explained by the iReady Diagnostic scores. The higher the iReady Diagnostic scores, the higher the 5th-grade TNReady ELA test scores. As a result of the analysis, the null hypothesis was rejected. The results are displayed in Table 22.

Table 22

Correlation Coefficients of iReady Diagnostic and TNReady ELA Scores for 5th-Grade

Category	*M*	*r*	*p*	r^2
iReady Diagnostic Scores	559.33	.78	.001	.61
TNReady ELA Test Scores	305.52			

Note. $p < .05$

Research Question 5a: Is there a significant relationship between Mock Interim scores and 5th-grade TNReady ELA test scores?

$H_0$5a: There is no significant relationship between Mock Interim scores and 5th-grade TNReady ELA test scores.

A Pearson correlation coefficient was calculated to determine the relationship between Mock Interim scores and 5th-grade TNReady ELA test scores. A significant, positive, strong correlation was found [r (353 df) = .70, p = <.05]. To determine the variance explained by Mock Interim scores on 5th-grade TNReady ELA test scores, a coefficient of determination was computed. The results (r^2= .49) suggest that 49% of the variance in 5th-grade TNReady ELA tests scores could be explained by the Mock Interim scores. The higher the Mock Interim scores, the higher the 5th-grade TNReady ELA test scores. As a result of the analysis, the null hypothesis was rejected. The results are displayed in Table 23.

Table 23

Correlation Coefficients of Mock Interim and TNReady ELA Scores for 5th-Grade

Category	*M*	*r*	*p*	r^2
Mock Interim Scores	40.92	.70	.001	.49
TNReady ELA Test Scores	305.52			

Note. $p < .05$

Research Question 5b: Is iReady Diagnostic Assessments or Mock Interim Assessments the best predictor of 5th-grade TCAP ELA test scores?

$H_0$5b: There is no best predictor of 5th-grade TCAP ELA test scores by iReady Diagnostic Assessments or Mock Interim Assessments.

Multiple linear regression was calculated to predict 5th-grade TCAP ELA test scores based on iReady Diagnostic and Mock Interim Assessment scores. A significant regression equation was found [$F(2, 350) = 347.975$, $p = .001$] with an R^2 of .665, this suggests that 66.5% of the variance in 5th-grade TCAP ELA test scores could be explained by the predictor variables. Results also suggest that 33.5% of the variance in 5th-grade TCAP ELA test scores could be explained by other variables other than the predictor variables. To determine which predictor variables were significant predictors of 5th-grade TCAP ELA test scores, *Beta* scores were examined. The iReady Diagnostic had a *Beta* score of .563, $p = .001$, and the Mock Interim had a *Beta* score of .321, $p = .001$. Both the iReady Diagnostic and Mock Interim Assessments were significant predictors of 5th-grade TNReady ELA test scores. However, the iReady Diagnostic was a stronger predictor than the Mock Interim Assessment of 5th-grade TNReady ELA test

scores. As a result of the analysis, the null hypothesis was rejected. The results are displayed in Table 24.

Table 24

Coefficients for Each Predictor Variable and the Dependent Variable for 5th-Grade ELA

Variable	*B*	*Beta*	*t*	*p*
iReady Diagnostic Assessment	.329	.563	13.311	.001*
Mock Interim Assessment	.594	.321	7.593	.001*

Note. *indicates significance at p <.05

Research Question 6

Research Question 6: Is there a significant relationship between iReady Diagnostic scores and 5th-grade TNReady math test scores?

$H_0$6: There is no significant relationship between iReady Diagnostic scores and 5th-grade TNReady math test scores.

A Pearson correlation coefficient was calculated to determine the relationship between iReady Diagnostic scores and 5th-grade TNReady math test scores. A significant, positive, strong correlation was found [r (360 df) = .81, p = <.05]. To determine the variance explained by iReady Diagnostic scores on 5th-grade TNReady math test scores, a coefficient of determination was computed. The results (r^2= .66) suggest that 66% of the variance in 5th-grade TNReady math tests scores could be explained by the iReady Diagnostic scores. The higher the iReady Diagnostic scores, the higher the 5th-grade TNReady ELA test scores. As a result of the analysis, the null hypothesis was rejected. The results are displayed in Table 25.

Table 25

Correlation Coefficients of iReady Diagnostic and TNReady Math Scores for 5th-Grade

Category	*M*	*r*	*p*	r^2
iReady Diagnostic Scores	473.47	.81	.001	.66
TNReady Math Test Scores	316.31			

Note. $p < .05$

Research Question 6a: Is there a significant relationship between Mock Interim scores and 5th-grade TNReady math test scores?

$H_0$6a: Is there a significant relationship between Mock Interim scores and 5th-grade TNReady math test scores?

A Pearson correlation coefficient was calculated to determine the relationship between Mock Interim scores and 5th-grade TNReady math test scores. A significant, positive, strong correlation was found [r (360 df) = .79, p = <.05]. To determine the variance explained by Mock Interim scores on 5th-grade TNReady math test scores, a coefficient of determination was computed. The results (r^2= .62) suggest that 62% of the variance in 5th-grade TNReady math tests scores could be explained by the Mock Interim scores. The higher the Mock Interim scores, the higher the 5th-grade TNReady math tests scores. As a result of the analysis, the null hypothesis was rejected. The results are displayed in Table 26.

Table 26

Correlation Coefficients of Mock Interim and TNReady Math Scores for 5th-Grade

Category	*M*	*r*	*p*	r^2
Mock Interim Scores	39.46	.79	.001	.62
TNReady Math Test Scores	316.31			

Note. $p < .05$

Research Question 6b: Is iReady Diagnostic Assessments or Mock Interim Assessments the best predictor of 5th-grade TCAP math test scores?

$H_0$6b: There is no best predictor of 5th-grade TCAP math test scores by iReady Diagnostic Assessments or Mock Interim Assessments.

Multiple linear regression was calculated to predict 5th-grade TCAP math test scores based on iReady Diagnostic and Mock Interim Assessment math scores. A significant regression equation was found [$F(2, 357) = 455.274$, $p = .001$] with an R^2 of .718, this suggests that 71.8% of the variance in 5th-grade TCAP math test scores could be explained by the predictor variables. Results also suggest that 28.2% of the variance in 5th-grade TCAP math test scores could be explained by other variables other than the predictor variables. To determine which predictor variables were significant predictors of 5th-grade TCAP math test scores, *Beta* scores were examined. The iReady Diagnostic had a *Beta* score of .510, $p = .001$, and the Mock Interim had a *Beta* score of .387, $p = .001$. Both the iReady Diagnostic and Mock Interim Assessments were significant predictors of 5th-grade TNReady math test scores. However, the iReady Diagnostic was a stronger predictor than the Mock Interim Assessment of 5th-grade TNReady math test

scores. As a result of the analysis, the null hypothesis was rejected. The results are displayed in Table 27.

Table 27

Coefficients for Each Predictor Variable and the Dependent Variable for 5th-Grade Math

Variable	*B*	*Beta*	*t*	*p*
iReady Diagnostic Assessment	.669	.510	11.366	.001*
Mock Interim Assessment	.921	.387	8.631	.001*

Note. *indicates significance at $p < .05$

Summary

In this chapter, the iReady Spring Diagnostic Assessments, the Tennessee Mock Interim Assessments, and the TCAP TNReady tests for ELA and math data from 995 students from seven elementary schools in Grades 3-5 from an East Tennessee school district were analyzed and presented. The students were administered the iReady Spring Diagnostic Assessments, the Tennessee Mock Interim Assessments, and TCAP TNReady tests for reading and math in the 2020-2021 school year. Six major research questions and their subset research questions and null hypotheses were addressed.

Results showed there was a significant relationship between the iReady Spring Diagnostic Assessment and the Tennessee Mock Interim Assessment scores on TNReady ELA and math test scores in Grades 3-5. The correlation coefficient for the iReady Spring Diagnostic Assessment scores showed a significant, positive, stronger correlation than the Tennessee Mock Interim Assessment scores on TNReady ELA test scores for Grades 3-5 and math for Grade 3

and Grade 5. The Tennessee Mock Interim scores showed a significant, positive, stronger correlation than the iReady Spring Diagnostic scores on TNReady Grade 4 math test scores.

Both the iReady Spring Diagnostic and the Tennessee Mock Interim Assessments were significant predictors of the TNCAP TNReady test scores in ELA and math for Grades 3-5. The iReady Spring Diagnostic Assessment was a stronger predictor than the Tennessee Mock Interim Assessments of TNCAP TNReady test scores in ELA for Grades 3-5 and math for Grade 3 and Grade 5. The Tennessee Mock Interim was a stronger predictor than the iReady Spring Diagnostic Assessment of TNCAP TNReady math test scores for Grade 4.

CHAPTER 5

Summary of Findings, Discussions, Conclusions, and Recommendations

This chapter contains a summary of findings, discussions, and conclusions. Additionally, this chapter includes recommendations for readers who may use the results of this research study to inform their school or school district's formative assessment practices and recommendations for further research. The purpose of this correlational research study was to examine the relationship between the scores of 3rd– 5th-grade students on iReady Spring Diagnostic for Tennessee (iReady Diagnostic) and the Tennessee Mock Interim Assessments (Mock Interim Assessment) and TCAP TNReady summative test scores for ELA and math. A related purpose of this study was to determine the predictive validity of iReady Diagnostic and Mock Interim Assessments scores for ELA and math on Grades 3-5 TCAP TNReady ELA and math performance.

Summary of Findings

The statistical analysis in this research study was based on six major research questions and their subset research questions and null hypotheses respectively. All six research questions yielded significant results and all null hypotheses were rejected. When each variable was measured independently of each other, both the iReady Diagnostic and the Mock Interim Assessments showed a significant positive relationship to TCAP TNReady test scores for ELA and math in Grades 3-5. The results showed there was a significantly stronger relationship between iReady Diagnostic to TNCAP TNReady test scores for ELA in Grades 3-5 and math in Grade 3 and Grade 5 compared to Mock Interim Assessment. However, the results showed there was a significantly stronger relationship between the Mock Interim Assessment and TNCAP TNReady test scores for Grade 4 math when compared to the iReady Diagnostic Assessment.

When the variables were measured together per grade level and subject, findings indicated that both the iReady Diagnostic and the Mock Interim Assessments were significant predictors of TCAP TNReady test scores for ELA and math. However, the iReady Diagnostic showed to be a stronger predictor than the Mock Interim Assessment for ELA in Grades 3-5 and math in Grade 3 and Grade 5. The Mock Interim Assessment showed to be a stronger predictor than iReady Diagnostic of TCAP TNReady test scores for Grade 4 math.

Discussion of Findings

The use of online formative computerized testing tools and systems is an increasingly common assessment practice in education due to their proven efficiency and flexibility. These testing tools enable districts, schools, and classrooms to manage, measure, predict, maximize student learning and achievement, and report real-time student data through analytics (Choi & McClenen, 2020). Due to modern accountability systems (Tennessee Department of Education, n.d.e), school leaders must locate and utilize the best formative computerized assessment tools and systems available that are aligned to state academic standards to measure the effectiveness of the teaching and learning cycle in preparation for high-stakes testing (Van der Kleij et al., 2015). Since the academic success of students, teachers, schools, and school systems are solely judged on high-stakes test scores, it is equally imperative that the formative assessment tools and systems used by schools and school districts are predictive of student performance on high-stakes summative tests (Immekus & Atitya, 2016). School district leaders must make the best financial decisions utilizing federal, state, and local funds for the success of their students' learning and achievement. More importantly, education leaders must select the best formative assessment programs for their respective student populations regarding functionality, validity, and reliability (Fitzpatrick et al., 2011; Herman, 2017). Education leaders must consider the

design of formative computerized assessment systems that will adequately fulfill the more important purpose first. Then they can consider if the same assessment can fulfill additional purposes or if multiple assessments are needed for a comprehensive assessment system (Perie et al., 2009).

The iReady Diagnostic, the Mock Interim, and the TCAP TNReady assessments provide criterion-referenced scores for various purposes related to the Tennessee Academic Standards. According to Curriculum Associates (2021a), iReady Diagnostic adaptive assessments were designed to provide schools and school districts with actionable data on student grade-level performance and proficiency levels based on a strong correlation to the Tennessee Academic Standards for ELA and math. By pinpointing student strengths and knowledge gaps, schools and school districts can differentiate instruction using targeted instructional tools. Curriculum Associates (2020, 2022) lists two Tennessee correlational studies linking iReady Diagnostic and TCAP TNReady scores for both ELA and math. According to Curriculum Associates (2020,2022), due to the strong correlation between iReady Diagnostic and TCAP TNReady test scores, the iReady Spring Diagnostic scores could be used to predict student performance levels on state summative tests.

The Mock Interim assessments were created to support educators in measuring how students are learning, progressing, performing, and mastering skills towards meeting grade-level expectations by pinpointing and identifying students' learning strengths and weaknesses and providing a pathway to adjust instructional practices to support ongoing learning (Tennessee Department of Education, 2020a; Tennessee Department of Education, 2020b). This was part of the Tennessee Department of Education's Reopening of Schools resources due to the COVID-19 pandemic school closures. According to the Tennessee Department of Education (2020b), the

Mock Interim Assessments "mirror the current TCAP summative assessments, as well as provide accurate scaled scores and performance bands that will estimate each student's performance" (p.12) on future TCAP TNReady tests via direct alignment to the Tennessee Academic Standards. The Tennessee Department of Education (2021d) states that the school districts that used the state department's free, formative assessment tools showed higher success rates.

The results of this study revealed that both the iReady Diagnostic and the Mock Interim assessment were significantly correlated to the TCAP TNReady tests for ELA and math. Additionally, the results of this study revealed that both the iReady Diagnostic and the Mock Interim Assessments were strong predictors of TCAP TNReady test performance in ELA and math. The findings of this research study are consistent with previous research studies in that the construct of formative computerized assessments must be aligned to state academic standards to attain a strong relationship and predictability of student performance on state summative exams (Bruce, 2019; Immekus & Atitya, 2016; Thomas, 2018).

However, the iReady Diagnostic assessments showed an overall stronger correlation than the Mock Interim assessments to the TCAP TNReady tests. The iReady Diagnostic assessments showed as an overall stronger predictor than the Mock Interim Assessments of student performance on TCAP TNReady tests for both ELA and math. These findings are consistent with previous correlational and predictive research studies regarding iReady Diagnostic assessments' correlation to state exams and predicting students' performance on state exams (Bruce 2019; Curriculum Associates, 2020, 2022; Thomas, 2018). This research study had overall lower correlation coefficients of .76 for ELA and .78 for math compared to the Curriculum Associates (2022) most recent correlation coefficients of .81 for ELA and .83 for math on the TCAP TNReady tests for Grades 3-5. Both this research study and the Curriculum

Associates 2021 linking study used the iReady Spring Diagnostic and the 2021 TCAP TNReady test scores. The difference in correlations between this study and the Curriculum Associates 2021 study when using iReady Spring Diagnostic could be attributed to the small sample that was used for this study.

Interestingly, this research study's ELA correlational average of .76 was closer to Curriculum Associates (2020) previous linking study that showed a correlational average of .77 for ELA, which was based on the 2016-2017 TCAP TNReady test scores. This could be explained by the small sample size of 8,000 students used by the Curriculum Associates for the 2017 study compared to the 2021 study conducted using a larger sample of 34,000 students. Regardless of the geographical differences, the small size sample used by the Curriculum Associates 2017 study was more comparable to the sample of this study.

Another reason why overall correlation coefficients for this study were lower than the Curriculum Associates 2021 study could be that during the 2020-2021 school year this East Tennessee school district offered virtual learning for all grades due to the COVID-19 pandemic. This was the school district's first experience with virtual teaching and learning for district leaders, teachers, and students. Virtual students were required to take all three assessments in person at their home school. Students who missed being at school could have been overly excited about returning to school on testing days affecting test performance. Conversely, students who did not want to be at school on testing days could have been less eager to perform well on the assessments. This could explain the differences in the correlation and predictive values found in this research study for the iReady Diagnostic, the Mock Interim Assessments, and TCAP TNReady test scores in comparison to the 2021 Curriculum Associates study.

Another reason why overall correlation coefficients for this study were lower than the Curriculum Associates 2021 study could be how the data were used from the assessments. Black & Wiliam (1998b) concluded that the use of formative assessment data in classrooms can raise student achievement, but “only by changes that are put into direct effect by teachers and pupils in the classrooms” (p. 148). Stiggins (2002) suggests for effective data use from formative assessments teachers must use descriptive feedback to build on prior knowledge and teachers must continuously adjust instruction to meet the needs of the students. Additionally, students must regularly engage in self-assessment and students need to recognize and understand the expected achievement goals and targets. This could explain the differences in the correlation and predictive values found in this research study for the iReady Diagnostic, the Mock Interim Assessments, and TCAP TNReady test scores in comparison to the Curriculum Associates 2021 study.

This research study also revealed that fourth-grade math was the only subject and grade for which the Mock Interim Assessment had a significantly stronger correlation and predictive value than the iReady Diagnostic to the TCAP TNReady test scores. This could be explained by the test questions that were used by the Mock Interim Assessment for fourth-grade math. The Mock Interim Assessment test questions “mirror the current TCAP summative assessments, as well as provide accurate scaled scores and performance bands that will estimate each student's performance" (Tennessee Department of Education 2020b, p.12) on future TCAP TNReady tests via direct alignment to the Tennessee Academic Standards. This could explain the differences in the correlation and predictive values found in this research study when comparing the iReady Diagnostic, the Mock Interim Assessments, and TCAP TNReady test scores for fourth-grade math.

Another reason why fourth-grade math was the only subject and grade for which the Mock Interim Assessment had a significantly stronger correlation and predictive value than the iReady Diagnostic to the TCAP TNReady test scores in this study could be explained by the assessments' computerized testing formats. The Mock Interim Assessment is based on a linear computerized based testing format that has a limited number of predetermined test questions based on that grade's academic standards. The iReady Diagnostic is based on a computerized adaptive testing format that differentiates test questions based on student responses that include a trajectory and a learning progression across a variety of grade-level standards. A limitation of computerized adaptive testing can be the test bank from which test items are selected if the size of the item test bank is not sufficient to cover a large range of skills taught over a range of difficulties, lack of up-to-date content test items (CSAI, 2019), or an algorithm issue with test question differentiation based on student responses (Thompson & Weiss, 2011). This could explain the differences in the correlation and predictive values found in this study when comparing the iReady Diagnostic, the Mock Interim Assessments, and TCAP TNReady test scores for fourth-grade math. The same was indicated in a previous correlation study comparing iReady Diagnostic test scores and Missouri Assessment Program (MAP) test scores in ELA and math for Grades 4-6 (Bruce, 2019). Bruce (2019) reported that fourth-grade math had the lowest significant correlation coefficient and predictive value to MAP test scores for both ELA and math. Interestingly, Curriculum Associates (2020, 2022) reported fourth-grade math as having the highest correlation in both Tennessee linking studies. However, the findings from this research study and that of Bruce (2019) suggest that Curriculum Associates may need to re-examine the fourth-grade math iReady Diagnostic assessment alignment to various state academic standards and state exams.

Lastly, the results in this study revealed that both the iReady Diagnostic and the Mock Interim Assessments were significantly correlated to and predictive of student performance on the TCAP TNReady tests for ELA and math. The overall results in this study showed the iReady Diagnostic when compared to the Mock Interim Assessments had a stronger correlation and predictive value to the TCAP TNReady test scores. The overall difference in the strength of significance in the correlation coefficients and predictive values between the iReady Diagnostic and the Mock Interim Assessments could be explained by the timing of the administration of the assessments. The Mock Interim Assessments were administered during January. The iReady Diagnostic assessments were administered from April 6 to May 14 with an extension no later than May 21 due to COVID-19 related guidelines and quarantines. The TCAP TNReady tests were administered to students between April 16 and May 6 of 2021. If a student was quarantined after completing a sub-part of the TCAP TNReady test, then make-up tests could be extended to June 10. The overall differences in the strength of correlations and predictive values between the iReady Spring Diagnostic and the Mock Interim Assessments to the TCAP TNReady test scores could be attributed to the intended purpose for which each computerized formative assessment was created and predetermined testing windows respectively.

Overall, the findings of this research study were that both formative computerized assessments produced the results according to their intended purposes and provided the school system with the data and information being sought.

Limitations of the Study

Results of this study provide useful information regarding the validity of formative computerized assessments. However, several limitations exist in the current research study. One possible limitation involves the administration of training and fidelity of administration of the

assessments. Although the training was provided to school personnel who were responsible for administering the assessments, the researcher did not observe the training and administration of the testing. Despite the state mandates requiring that certified personnel were to administer respective assessments, the researcher did not observe or verify that certified personnel were the only ones to administer the assessments.

A second limitation was the variations among the assessments. The iReady Diagnostic and the Mock Interim Assessments are not timed tests as opposed to the TCAP TNReady tests, which adhere to stringent timeframes. Additionally, the iReady Diagnostic is a computerized formative adaptive assessment with testing times based on individual student responses. As students progress on the iReady Diagnostic, the test becomes increasingly more difficult and testing time can increase. This could impact the results as students begin to tire. The Mock Interim Assessment is a computerized linear formative assessment that has a limited number of items per test. This causes inconsistent testing procedures.

Another limitation was that administrators and teachers had the option of giving the Mock Interim Assessments via pencil-paper. If the Mock Interim Assessments were administered via pencil-paper, then the tests had to be scored and entered manually in Schoolnet platform. Additionally, if students took the pencil-paper version, this could have afforded them an opportunity to revisit and change answers whereas on the iReady Diagnostic this cannot be done. The researcher did not observe the fidelity of pencil-paper administration, manual scoring, and entering scores.

Conclusions

This research study shows that computerized formative assessments must be constructed and closely aligned with state academic standards to be deemed a strong valid predictor of

student performance levels on summative high-stakes tests. The findings suggest that although the iReady Spring Diagnostic and the Mock Interim Assessments are online formative assessments with different computer-based testing formats, they are both closely aligned and show a relationship of predictability of student performance levels on the TCAP TNReady tests. This shows why it is important that school leaders should take into consideration the design of online formative assessment tools and assessment systems that will adequately fulfill their more important purpose first. Then they can consider if the same assessment can fulfill additional purposes or if multiple assessments are needed for a comprehensive assessment system (Perie et al., 2009; Perie et al. 2007) to meet the current accountability measures placed on school districts and schools via student high-stakes test scores.

The findings in this study support the use of technology-based online formative assessment tools and systems for instant accessibility of student data to provide educators with timely and adequate student data for effective decision making at the classroom, school, and district levels (Choi & McClenen, 2020) such as the iReady Diagnostic and the Mock Interim Assessments. For Tennessee educators, the reliance on using computerized formative assessment data effectively continues to mount in light of key pieces of Tennessee Legislation such as the *Tennessee Learning Loss And Student Acceleration Act (SB7002/HB7004)* and the *Tennessee Literacy Success Act (SB7003/HB7002)* as well as Tennessee teacher evaluation procedures and practices. Therefore, education leaders should understand the rationale behind using computerized formative assessments to ensure that the data from formative assessments are used effectively and efficiently at the district, school, and classroom levels. Black and Wiliam (1998b) stated that the use of formative assessment data in classrooms can raise student achievement, but "only by changes that are put into direct effect by teachers and pupils in the classrooms" (p.

148). If school districts and schools use formative computerized assessments and are not providing ongoing teacher support for effective data use, then they should consider it. This could also provide an extra layer of support for educators that could help reduce teacher attrition.

Additionally, this study shows that some computerized adaptive assessment systems can serve multiple purposes such as the iReady Diagnostic assessment system. The overall findings in this study suggest and support that some computerized adaptive assessment systems can be utilized and function as a diagnostic screening tool to meet federal and state government mandates for RTI^2 and as a formative computerized adaptive assessment system to assess student learning and progress toward meeting state academic standards as measured on state summative standardized tests. If school districts use computerized formative adaptive assessments to pinpoint student learning on a specified learning progression and identify students at risk of not meeting grade-level proficiency on state summative tests, then this allows educators time to intervene and adjust instruction before state summative testing (Bennett. 2015; Bruce, 2019; Collares & Cecilio-Fernandes, 2019; Immekus & Atitya, 2016; Shapiro et al., 2015; Thompson, 2018; Van der Kleij et al., 2015).

Lastly, "One of the most neglected aspects of CAT is its alignment with modern learning theories" (Collares & Cecilio-Fernandes, 2019, p. 115). The results and findings of this research study provide strong evidence that formative computerized adaptive assessments, such as iReady Diagnostic, can serve as a formative assessment approach. They can also serve as effective decision-making tools to provide educators access to evidence of student learning, achievement, mastery of skills, and predictive purposes of student performance on state tests as well as provide evidence of an assessment system's effectiveness to fulfill multiple purposes. The "overarching formative assessment and formative evaluation approach could lead to more valid decisions at

different levels in schools … in a thoughtful way can lead to more valid formative decisions" (Van der Kleij et al., 2015, p. 15) by education leaders and educators. If education leaders do not have an understanding of how various learning theories show relevance at different stages of learning (Van der Kleij et al., 2015), then formative assessment data cannot be used fully to maximize the interrelated complexity of teaching, learning, and assessments. This can support education leaders in making the best decisions to identify, select, and use the right type of formative computerized assessment that will provide support to schools, teachers, and students for ongoing student growth and achievement for their respective student populations regarding functionality, validity, and reliability as measured by the state summative test scores.

Recommendations for Practice

The results of this study support and show the importance of the construct and alignment of formative computerized assessments to state academic standards to attain the strongest relationship of predictability of student performance levels on state summative tests. It is recommended that school leaders adopt formative computerized assessments that are highly aligned and correlated to state summative tests. This is necessary for formative computerized assessments to function as a reliable means of measurement of student learning and for student learning to improve student performance levels toward meeting the goals of state academic standards assessed on state exams annually beforehand (Bruce, 2019; Immekus & Atitya, 2016; Thomas, 2018).

Based on the results and findings in this study the iReady Diagnostic system and the Mock Interim Assessments for ELA and math in Grades 3-5 should be considered by Tennessee school leaders as viable formative computerized assessment options that can be used to improve student performance levels on TNCAP TNReady tests. If the iReady Diagnostic Assessment

system and Mock Interim Assessments are selected by education leaders as formative computerized assessment options, then they should provide teachers with high-quality professional learning to use formative assessment data effectively and efficiently in the classroom (Black & Wiliam, 1998b; Marshall, 2008; Perie et al., 2009; Stiggins, 2002, 2005). This ensures students are receiving the targeted instruction they need to be successful learners. Lastly, education leaders should provide educators with multiple layers of support throughout the school year surrounding the use of data from formative computerized assessments considering the demands placed on educators within the current accountability systems that place a great emphasis on student high-stakes test scores (Ingersoll, 2001, 2012; Ingersoll & Smith, 2004; Norton, 2015; Stiggins, 2005). This ensures that teachers are receiving the support needed for their next steps regarding instructional decisions they need to implement to be successful educators.

Based on the overall findings of this study, education leaders should identify the goal of the formative computerized assessment's intended purposes and ensure that the assessments will meet the criteria of intended purposes first (Fitzpatrick et al., 2011; Herman, 2017; Perie et al., 2009). Then they can consider if the same assessment can fulfill additional purposes or if multiple assessments are needed for a comprehensive assessment system (Perie et al., 2009). When purchasing or selecting formative computerized assessment systems, such as iReady Diagnostic Assessments or Tennessee Mock Interim Assessments, school district leaders should develop a theory of action based on the expectations of the product or program to improve student learning and achievement (Herman, 2017). School leaders should demand evidence to evaluate the quality of formative computerized assessments based on the following criteria: alignment, diagnostic value, fairness, technical quality, utility, and feasibility (Herman, 2017).

There are important characteristics to consider when making purchasing decisions or assessment selections and school leaders should demand evidence of each. Lastly, school districts and school leaders should ensure that the product or program evaluation process for formative computerized assessment tools and systems is rigorous, systematic, and ethical (Fitzpatrick et al., 2011).

Additionally, the online formative computerized assessments selected by school leaders should provide a continuous stream of real-time data that will accurately identify and pinpoint students' learning on a specified learning progression and identify students at risk of not meeting grade-level proficiency on state summative tests. This will allow educators time to intervene and adjust instruction before state summative testing (Bennett. 2015; Bruce, 2019; Collares & Cecilio-Fernandes, 2019; Immekus & Atitya, 2016; Shapiro et al., 2015; Thompson, 2018; Van der Kleij et al., 2015). If school districts and schools use the iReady Diagnostic assessment and the Mock Interim Assessments for ELA and math in Grades 3-5, they should use them according to their intended purposes.

School systems, school leaders, and teachers should use the data from the Mock Interim Assessments given in January to monitor student progress and identify student strengths and weaknesses based on academic standards and skills taught. Then, based on the Mock Interim's predictive value of student performance on the TCAP TNReady tests, educators should monitor and adjust instruction accordingly to continue to support student learning, growth, and achievement toward meeting state proficiency levels. Then approximately two to three weeks before administering the TNCAP TNReady tests, the iReady Spring Diagnostic should be given. iReady's diagnostic data should be used by educators to pinpoint where students are on their learning trajectories and progressions to monitor and adjust instruction accordingly to meet student needs. iReady's predictive data of student performance on the TCAP TNReady tests

should be used by educators to hone in on specific standards and skills and adjust instruction accordingly to support, promote, and enhance student performance on TCAP TNReady tests. In addition, this affords education leaders to use the iReady Diagnostic assessment system in multiple capacities as the iReady Spring Diagnostic testing window meets the federal and state mandates for universal screeners and RTI^2.

Recommendations for Further Study

The results of this study indicate that iReady Spring Diagnostic and the Mock Interim Assessments are significantly correlated to and predictors of student performance on TCAP TNReady tests for ELA and math in Grades 3-5. The results of this study also indicate that the iReady Spring Diagnostic had an overall stronger correlation and predictor values than the Mock Interim assessments to the TCAP TNReady tests for ELA and math in Grades 3-5. However, the overall findings of this research study were that both formative computerized assessments produced the results according to their intended purposes. A replication of this study should be conducted in similar schools and school districts since formative computerized assessments are increasingly used for monitoring and measuring student progress and achievement and guiding instructional decisions making it crucial for researchers to examine the validity of assessment instruments currently used. An expansion of this study which would include subgroups, such as gender, ethnicity, English language proficiency, low-income status, and RTI^2 students, would be beneficial. Additionally, a comparative study is recommended using other formative computerized assessments that are presently used throughout school districts such as AIMSWeb and MasteryConnect. Also, a comparative study should be conducted using formative computerized adaptive assessments as universal screeners and predictors of student performance on TCAP TNReady tests across multiple grade bands. Finally, it is recommended that a

qualitative research study be conducted to gain information and insight on how formative computerized assessment data are used by educators at the district level, school level, and classroom level.

References

Aguilar, R. E. (2019). *Predictive validity of i-Ready diagnostic scores for high-stakes test outcomes with middle school students: A program evaluation* (Publication No. 22587119) [Doctoral dissertation, University at Albany]. ProQuest LLC.

Balakit, M. (2016, April 15). Tennessee phases out common core. *Tennessean*. https://www.tennessean.com/story/news/education/2016/04/15/tennessee-phases-out-common-core/83085942/

Bennett, R. E. (2015). The Changing Nature of Educational Assessment. *Review of Research in Education*, 39, 370–407. http://www.jstor.org/stable/44668662

Bernstein, L. (2021, June 11). COVID-19 relief: Everything you need to know about ESSER funding. *Xello Blog*. https://xello.world/en/blog/esser-funding/

Birman, B. F., & Porter, A. C. (2002). Evaluating the effectiveness of education funding streams. *Peabody Journal of Education*, 77(4), 59-85). https://www.jstor.org/stable/1493218

Bjorklund-Young, A., & Borkoski, C. (2016). Do formative assessments influence student learning: Research on iReady and MAP. *Institute for Education Policy: Johns Hopkins School of Education*. https://edpolicy.education.jhu.edu/formative-assessments/

Black, P., & Wiliam, D. (1998a). Assessment and classroom learning. *Assessment in Education: Principles, Policy, & Practice*, 5(1), 7-74. http://dx.doi.org/10.1080/0969595980050102

Black, P., & Wiliam, D. (1998b). Inside the black box: Raising standards through classroom assessment. *The Phi Delta Kappan*, 80(2), 139–148.

Bruce, K. D. (2019). *Online formative assessments as a predictor of state summative scores* (Publication No. 27671198) [Doctoral dissertation, William Woods University]. ProQuest LLC.

Bulkley, K. E., & Burch, P. (2011). The changing nature of private engagement in public education: For-Profit and nonprofit organizations and educational reform. *Peabody Journal of Education*, 86(3), 236–251. http://www.jstor.org/stable/23048626

Center on Standards and Assessments Implementation (CSAI). (2019, March 31). *Implementation of computerized adaptive assessments* [CSAI brief]. WestEd Center on Standards and Assessments Implementation. https://eric.ed.gov/?id=ED603014

Choi, Y., & McClenen, C. (2020, November). Development of adaptive formative assessment system using computerized adaptive testing and dynamic bayesian networks. *Applied Sciences*. http://dx.doi.org/10.3390/app10228196

Collares, C. F., & Cecilio-Fernandes, D. (2019). When I say … computerised adaptive testing. *Medical Education*, 53(2), 115–116. https://doi.org/10.1111/medu.13648

Croteau, J.L. (2011). *Online formative assessments as predictors of student academic success* (Publication No. 1145) [Doctoral dissertation, Walden University]. Walden Dissertations and Doctoral Studies.

Curriculum Associates. (2020). *i-Ready diagnostic: Linking study with TNReady*. https://www.curriculumassociates.com/-/media/mainsite/files/i-ready/iready-diagnostic-assessments-linking-study-overview-tennessee-2020.pdf

Curriculum Associates. (2021a). *iReady assessment: Tennessee*. https://www.curriculumassociates.com/products/i-ready/i-ready-assessment

Curriculum Associates. (2021b). *iReady: Tennessee*. https://www.curriculumassociates.com/

Curriculum Associates. (2022, February). *i-Ready diagnostic: Correlation study with Tennessee Comprehensive Assessment Program (TCAP)*. https://www.curriculumassociates.com/-

/media/mainsite/files/i-ready/iready-diagnostic-assessments-correlation-study-tennessee-2022.pdf

Dunn, K. E., & Mulvenon, S. W. (2009). A critical review of research on formative assessment: The limited scientific evidence of the impact of formative assessment in education. *Practical Assessment, Research & Evaluation*, 14(7), 1–11.

Filsecher, M. & Kerres, M. (2012). Repositioning formative assessment from an educational assessment perspective: A response to Dunn & Mulvenon (2009*). Practical Assessment, Research, and Evaluation*. 17(16), 1-9.

Fitzpatrick, J. L., Sanders, J. R., & Worthen, B. R. (2011). *Program evaluation: Alternative approaches and practical guidelines*. Upper Saddle River, NJ: Pearson Education.

Great Schools Partnership. (2014, August 18). High-stakes test. In *The Glossary of Educational Reform*. Retrieved January 1, 2022, from https://www.edglossary.org/high-stakes-testing/

Hall, P., & Simeral, A. (2008). *Building teachers' capacity for success: A collaborative approach for coaches and school leaders.* ASCD.

Herman, J. (2017). Interim assessments in brief. *WestEd Center on Standards and Assessment Implementation* (CSAI). https://csaa.wested.org/resource/interim-assessments-in-brief

Hornbeck, D. (2017, April 26). Federal role in education has long history. *The Conservation*. https://theconversation.com/federal-role-in-education-has-a-long-history-74807

Immekus, J. C., & Atitya, B. (2016). The predictive validity of interim assessment scores based on the Full-Information Bifactor Model for the prediction of end-of-grade test performance. *Educational Assessment*, 21(3), 176–195. http://dx.doi.org/10.1080/10627197.2016.120210

Ingersoll, R (2001). Teacher turnover and teacher shortages: An organizational analysis. *American Educational Research Journal*, 38(3), 499-534. https://www.academia.edu/2464657/Teacher_turnover_and_teacher_shortages_An_organizational_analysis

Ingersoll, R. M. (2012). Beginning teacher induction: WHAT THE DATA TELL US. *The Phi Delta Kappan*, 93(8), 47–51. https://eds-b-ebscohost-com.milligan.idm.oclc.org/eds/pdfviewer/pdfviewer?vid=1&sid=f7fc9edc-7ddb-4e0d-a67c-f637640f6188%40sessionmgr103

Ingersoll, R., Merrill, E., Stuckey, D., & Collins, G. (2018). *Seven trends: The transformation of the teaching force--updated October 2018*. https://repository.upenn.edu/cpre_researchreports/108

Ingersoll, R., & Smith, T. M. (2004). *Do teacher induction and mentoring matter?*. https://repository.upenn.edu/cgi/viewcontent.cgi?article=1134&context=gse_pubs

i-Ready Central. (2022*). i-Ready reference sheet diagnostic results (Students)*. Curriculum Associates. https://i-readycentral.com/download/?res=15519&view_pdf=1

Kornhaber, M. L., Barkauskas, N. J., & Griffith, K. M. (2016). Smart money? Philanthropic and federal funding for the common core. *Education Policy Analysis Archives*, *24*(93), 1-34.

Lang, J., & Tay, L. (2020, September). The science and practice of Item Response Theory in organizations. *Annual Review of Organizational Psychology and Organizational Behavior*, 8(1), 311-338. https://doi.org/10.1146/annurev-orgpsych-012420-061705

Marshall, K. (2008). Interim assessments: A user's guide. *The Phi Delta Kappan*, 90(1), 64–68.

McGuinn, P. (2015). Schooling the State: ESEA and the Evolution of the U.S. Department of Education. *RSF: The Russell Sage Foundation Journal of the Social Sciences*, 1(3), 77–94. https://doi.org/10.7758/rsf.2015.1.3.04

McShane, M. Q. (2021). *Where conservatives should lead on federal education policy in 2021*. In AEI Paper & Studies (pp. 1-9). American Enterprise Institute. https://link.gale.com/apps/doc/A652078851/AONE?u=tel_a_welshimer&sid=bookmark-AONE&xid=ebf2e91e

Muijs, D. (2011). *Doing quantitative research in education with SPSS* (Second Edition). Sage.

Neligan, B. (2020). *Moving towards computer adaptive testing: The effect of experience with technology on elementary students' scores and attitudes* (Publication No. 85) [Doctoral dissertation, Saint John's University]. St. John's University Theses and Dissertations.

Nagandla, K., Sulaiha, S., & Nalliah, S. (2018). Online formative assessments: Exploring their educational value. *Journal of Advances in Medical Education & Professionalism*, 6(2), 51–57.

National Council of Teachers of English. (2014, November). *How standardized tests shape- and limit - student learning*. https://library.ncte.org/journals/CC/issues/v24-2/26173

Norton, M. S. (2015). *Principal as human resources leader: A guide to exemplary practices for personnel administration.* Routledge.

Onosko, J. (2011). Race to the top leaves children and future citizens behind: The devastating effects of centralization, standardization, and high stakes accountability. *Democracy & Education*, 19(2), 1-12. https://eds-p-ebscohost-com.milligan.idm.oclc.org/eds/pdfviewer/pdfviewer?vid=1&sid=69fbecb2-1d7b-4a21-be33-23c6b34a2c62%40redis

Ornstein, A. C. (1984). The changing federal role in education. *American Education*, *20*, 1-3. https://link.gale.com/apps/doc/A3535942/AONE?u=tel_a_welshimer&sid=bookmark-AONE&xid=3c01fcb6

Pearson. (2015). *Schoolnet overview brochure*. https://www.pearsonassessments.com/content/dam/school/global/clinical/us/assets/schoolnet/schoolnet-overview-brochure.pdf

Perie, M., Marion, S., & Gong, B. (2009). Moving toward a comprehensive assessment system: A framework for considering interim assessments. *Educational Measurement: Issues and Practice*, 28(3), 5–13. http://dx.doi.org.milligan.idm.oclc.org/10.1111/j.1745-3992.2009.00149.x

Perie, M., Marion, S., Gong, B., Wurtzel, J., Aspen Institute, Achieve, I., & National Center for the Improvement of Educational Assessment, I. (NCIEA). (2007). *The role of interim assessments in a comprehensive assessment system: A policy brief*. In Aspen Institute. Aspen Institute. https://files-eric-ed-gov.milligan.idm.oclc.org/fulltext/ED551318.pdf

Popham, W. J. (2006a). *Defining and enhancing formative assessment*. [Paper presented]. October 10-13, 2006 CCSSO State Collaborative on Assessment and Student Standards FAST meeting, Austin, TX. Microsoft Word - Popham 9-15-06 DEFINING FORMATIVE ASSESSMENT version 2.0.doc (brjonesphd.com)

Popham, W. J. (2006b). Phony formative assessments: Buyer beware!. *Educational Leadership*, 64 (3), 86-87.

Ronfeldt, M. & McQueen, K. (2017). Does new teacher induction really improve retention? *Journal of Teacher Education*, 68(4), 394-410. http://dx.doi.org.milligan.idm.oclc.org/10.1177/0022487117702583

Shapiro, E. S., Dennis, M. S., & Fu, Q. (2015). Comparing computer adaptive and curriculum-based measures of math in progress monitoring. *School Psychology Quarterly*, 30(4), 470-487

Shepard, L. (2016). Testing and assessment for the good of education: Contributions of AERA presidents, 1915-2015. *Educational Researcher*,45(2), 112-121. http://www.jstor.org/stable/43996906

State Collaboration on Reforming Education (SCORE). (n.d.) *Seizing the moment: 2022 State of education in Tennessee*. https://tnscore.org/wp-content/uploads/2021/12/2022-State-Of-Education_digital_FINAL.pdf

Stiggins, R. J. (2002). Assessment crisis: The absence of assessment FOR learning. *Phi Delta Kappan*, 83(10), 758–765.

Stiggins, R. J. (2005). From formative assessment to assessment for learning: A path to success in standards-based schools. *Phi Delta Kappan*, 87(4).

Stiggins, R., & DuFour, R. (2009). Maximizing the power of formative assessments. *The Phi Delta Kappan*, *90*(9), 640–644.

Styron, J., & Styron, R. (2012). Teaching to the test: A controversial issue in quantitative measurement. *Systemics, Cybernetics, and Informatics*, 10(5), 22-25.

Sutter, C.C., Campbel, L.O., & Lambie, G. W. (2020). Predicting second-grade students' yearly standardized reading achievement using a computer-adaptive assessment. *Computers in the Schools*, 37(1), 40–54. https://doi-org.milligan.idm.oclc.org/10.1080/07380569.2020.1720611

Tatter, G. (2015, December 15). *As Tennessee finishes its race to the top, teachers caught in the middle of competing changes*. Chalkbeat.

https://tn.chalkbeat.org/2015/12/15/21098672/as-tennessee-finishes-its-race-to-the-top-teachers-caught-in-the-middle-of-competing-changes

Tennessee Department of Education. (n.d.a). *2021 Special session legislation*. https://www.tn.gov/education/top-links/2021-special-session-legislation.html

Tennessee Department of Education. (n.d.b). *How TNReady is created for Tennessee students?*. TNReady: Resources. https://www.tn.gov/content/dam/tn/education/documents/How_TNReady_is_Created_FINAL.pdf

Tennessee Department of Education. (n.d.c). *Learning loss in Tennessee*. 2021 Special session legislation: Tennessee Learning Loss Remediation and Student Acceleration Act (SB7002/HB7004). https://www.tn.gov/content/dam/tn/education/2020-21-leg-session/Overall%20Tennessee%20Learning%20Loss%20Remediation%20&%20Student%20Acceleration%20Act.pdf

Tennessee Department of Education. (n.d.d). *Literacy in Tennessee*. 2021 Special session legislation: Literacy Success Act (SB7003/HB7002). https://www.tn.gov/content/dam/tn/education/2020-21-leg-session/Overall%20Tennessee%20Literacy%20Success%20Act.pdf

Tennessee Department of Education. (n.d.e). *Overview of testing in Tennessee*. https://www.tn.gov/education/assessment/testing-overview.html

Tennessee Department of Education. (n.d.f). *TNReady Score Reports*. https://www.tn.gov/education/assessment/tnready/tnready-score-reports.html

Tennessee Department of Education. (2018, August 13). *Every Student Succeeds Act: Building on success in Tennessee*. Tennessee's ESSA Plan.

https://www.tn.gov/content/dam/tn/education/documents/TN_ESSA_State_Plan_Approved.pdf

Tennessee Department of Education. (2020a, June 5). *TDOE Announces innovative assessment supports for 2020-21* [Press Release]. https://www.tn.gov/education/news/2020/6/5/tdoe-announces-innovative-assessment-supports-for-2020-21.html

Tennessee Department of Education. (2020b, June 11). *School reopening kit: Assessing student learning.* https://www.tn.gov/content/dam/tn/education/health-&-safety/Assessing%20Student%20Learning.pdf

Tennessee Department of Education. (2021a, January 5). *Tennessee universal reading screener administration guidelines.* 2021 Special session legislation: Literacy Success Act (SB7003/HB7002). https://www.tn.gov/content/dam/tn/education/2020-21-leg-session/URS%20Administration%20Guidelines%20Final.pdf

Tennessee Department of Education. (2021b, February 1*). Learning loss: Summary of Tennessee Learning Loss and Student Acceleration Act for districts and school boards*. 2021 Special Session Legislation: Tennessee Learning Loss Remediation and Student Acceleration Act (SB7002/HB7004). https://www.tn.gov/content/dam/tn/education/2020-21-leg-session/Learning%20Loss%20District%20and%20School%20Board%20Summary%20of%20Legislation%20-%20508.pdf

Tennessee Department of Education. (2021c, May 19). *TDOE announces state spending strategy for nearly $4.2 billion in historic federal stimulus funding for K-12 education* [Press Release].

https://www.tn.gov/education/news/2021/5/19/tdoe-announces-state-spending-strategy-for-nearly--4-2-billion-in-historic-federal-stimulus-funding-for-k-12-education.html

Tennessee Department of Education. (2021d, August 2). *Tennessee releases 2020-21 spring TCAP state-level results* [Press Release]. https://www.tn.gov/education/news/2021/8/2/tennessee-releases-2020-21-spring-tcap-state-level-results-.html

Tennessee State Government. (2010a, January 26). *Bredesen signs landmark education bills into law* [Press Release]. https://www.tn.gov/news/2010/1/26/bredesen-signs-landmark-education-bills-into-law.html

Tennessee State Government. (2010b, March 29). *Tennessee wins race to the top grant* [Press Release]. https://www.tn.gov/news/2010/3/29/tennessee-wins-race-to-the-top-grant.html

The Institute of Education Sciences (IES). (2009, September). *Using student achievement data to support instructional decision making.* https://ies.ed.gov/ncee/wwc/Docs/PracticeGuide/dddm_pg_092909.pdf

The Tennessee Public Education Coalition. (2021, March 20). Tennessee students need fully funded literacy program: Opinion. *Tennessean*. https://www.tennessean.com/story/opinion/2021/03/19/struggling-tennessee-students-need-fully-funded-literacy-program/4756760001/

Thompson, H. A. (2018). *The relationship between i-Ready Diagnostic and 10th grade students' high-stakes mathematics test scores* (Publication No. 5042) [Doctoral dissertation, Walden University]. Walden Dissertations and Doctoral Studies.

Thompson, N. A., & Weiss, D. J. (2011). A framework for the development of computerized adaptive tests. Practical Assessment, Research & Evaluation, 16(1), 1-10. https://search-

ebscohostcom.milligan.idm.oclc.org/login.aspx?direct=true&db=eric&AN=EJ933695&site=eds-live&scope=site

Torres, R. A. (2019). *The effect of the I-ready reading program on student scores on the Northwest Evaluation Association (NWEA®) Measures of Academic Progress (MAP) Reading Assessment* (Publication No. 1167) [Doctoral dissertation, Cleveland State University]. Cleveland State University ETD Archive.

United Nations Educational, Scientific and Cultural Organization (UNESCO) (2021, March 25). *100 million more children under the minimum reading proficiency level due to the Covid-19--UNESCO convenes world education ministers* [Press Release]. https://en.unesco.org/news/100-million-more-children-under-minimum-reading-proficiency-level-due-covid-19-unesco-convenes

U.S. Department of Education. (2016, October). *Race to the top: Implementation and relationship to student outcomes executive summary*. [Institute of Education Sciences (IES)]. https://ies.ed.gov/ncee/pubs/20174001/pdf/20174000.pdf

United States Department of Education. (2004). *No child left behind (NCLB)*. https://www2.ed.gov/nclb/overview/intro/execsumm.html

United States Department of Education. (2010). *A blueprint for reform: The reauthorization of the elementary and secondary education act*. ERIC. hhttps://files.eric.ed.gov/fulltext/ED508795.pdf

United States Department of Education. (2015). *Every student succeeds act (ESSA)*. https://www.ed.gov/essa

Van der Kleij, F., Vermeulen, J., Schildkamp, K., & Eggen, T. (2015) Integrating data-based decision making, assessment for learning and diagnostic testing in formative assessment.

Assessment in Education Principles Policy and Practice, 22(3):1-20. http://dx.doi.org/10.1080/0969594X.2014.999024

Vygotsky, L.W. (1978). *Mind in society: The development of higher psychological processes* (Cole, M., John-Steiner, V., Scribner, S., & Souberman, Trans.). London: Harvard University Press.

Webb, J.W. (2005). *An analysis of Tennessee Gateway Exams and the variables related to student results* (Publication No. 1064) [Doctoral dissertation, East Tennessee State University]. Educational Assessment, Evaluation, and Research Commons.

Wiliam, D. (2014, April). Formative assessment and contingency in the regulation of learning processes. *Toward a theory of classroom assessment as the regulation of learning* [Paper Presented] American Educational Research Association, Philadelphia, PA. https://famemichigan.org/wp-content/uploads/2018/06/Wiliam-Formative-assessment-and-contingency-in-the-regulation-of-learning-processes.pdf

Zanon, C., Hutz, C., Yoo, H., & Hambleton, R. (2016, April 4). An application of item response theory to psychological test development. *Psicologia: ReflexÃ£o e CrÃ-tica*, 29:18(1), 1-10. https://doi.org/10.1186/s41155-016-0040-x

Zhou, M., & Brown, D. (2015). *Educational learning theories*. Education Open Textbooks. https://oer.galileo.usg.edu/education-textbooks/1

APPENDIXES

Appendix A

School District Letter: Permission to Conduct Research Study

Permission to Conduct Research Study

To: _____, Director of Schools

_____ Schools

FROM: Amy A. McAmis

DATE: November 9, 2021

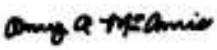
SUBJECT: District Permission to Conduct Research Study

As a graduate student at Milligan University, I am conducting a research study as part of fulfilling the requirements for a doctoral degree in Educational Leadership. The title of my research project is *The Relationship Between iReady Diagnostic and Mock Interim Assessments on Grades 3-5 Student Performance on TCAP Scores*. The purpose of this comparative quantitative study is to examine the relationshipp between using iReady Diagnostic and the Mock Interim Assessments for ELA and math on student performance levels on TCAP scores in Grades 3-5. The second purpose of the study is to determine the predictive validity of iReady Diagnostic Assessments and Mock Interim Assessments scores for ELA and math on Grades 3-5 TCAP ELA and math performance. Examining the relationship between these formative assessments and student performance levels on TCAP scores and their predictive validity could assist district personnel's decision-making to select the better formative assessment tools aligned to state academic standards as a means of measurement of student learning and achievement to improve student performance levels to meet the goals of the state academic standards assessed on TCAP tests.

I am writing to request your permission to conduct this research study in _____ Schools and approval to use archival testing data from iReady Diagnostic for Tennessee, the Tennessee Mock Interim Assessments, and TCAP scores for reading and math for students who were in Grades 3 through Grade 5 for the 2020-2021 school year. There will not be any student identifiable data used nor schools referenced in this study. _____, District Data Supervisor, has graciously agreed to serve as liaison in the data collection process to assure the confidentiality of students and the school district.

I will be conducting this study under the supervision of Milligan University. I believe the results of the study will enable our school district to better serve our students. I appreciate your consideration of this important matter. Thank you!

Respectfully,



Amy A. McAmis

Appendix B

Milligan University Institutional Review Board Approval Letter

MILLIGAN
UNIVERSITY

Date: November 17, 2021

Principal Investigator: **Amy McAmis**, Graduate Student, Milligan University
From: The Institutional Review Board (IRB) at Milligan University
Project: *A Comparative Study of Using iReady Diagnostic and Mock Interim Assessments on Grades 3-5 Student Performance on TCAP Scores*
IRB Tracking Number: **2021-15**
IRB Approval Number: **Exp2111171503**
Subject: **Final Approval**

On behalf of the Milligan University Institutional Review Board (IRB), we are writing to inform you that the above-mentioned study has been approved as expedited. This approval also indicates that you have fulfilled the IRB requirements for Milligan University.

All research must be conducted in accordance with this approved submission, meaning that you will follow the research plan you have outlined here, use approved materials, and follow university policies.

Take special note of the following important aspects of your approval:

- Any changes made to your study require approval from the IRB Committee before they can be implemented as part of your study. Contact the IRB Committee at **IRB@milligan.edu** with your questions and/or proposed modifications;
- If there are any unanticipated problems or complaints from participants during your data collection, you must notify the Milligan University IRB Office within 24 hours of the data collection problem or complaint;
- Milligan University requires specific formatting when collecting demographic data on gender; please contact me if you need assistance with this formatting.

The Milligan University IRB Committee is pleased to congratulate you on the approval of your research proposal. Best wishes as you conduct your research! If you have any questions about your IRB Approval, please contact the IRB Office and copy your faculty advisor if appropriate on the communication.

On behalf of the IRB Committee,

Trini Rangel, Ph.D.
Chair, Institutional Review Board
Milligan University

Milligan, TN 37682 | 423.461.8700 | www.milligan.edu