5-2018

# TCAP Assessment in Correlation with and as Compared by STAR Assessment

Brooke Sampson
*East Tennessee State University*

TCAP Assessment in Correlation with and as Compared by STAR Assessment

---

A dissertation

Presented to

The faculty of the Department of Educational Leadership and Policy Analysis

East Tennessee State University

In partial fulfillment

Of the requirement for the degree

Doctor of Education of Educational Leadership

---

by

Brooke Sampson

Spring 2018

---

Dr. William Flora, Chair

Dr. Don Good

Dr. Pamela Scott

Dr. Stephanie Tweed

Keywords: STAR assessment, TCAP assessment, assessment correlation, standardized assessment, math, reading

ABSTRACT

TCAP Assessment in Correlation with and as Compared by STAR Assessment

by

Brooke Sampson

The purpose of the study was twofold. The first purpose of the study was to determine if a correlation existed between the Standardized Test for the Assessment of Reading (STAR), created and distributed by Renaissance, and the Tennessee Comprehensive Assessment Program (TCAP) Achievement Test in Math and Reading for grade 3, grade 4, and grade 5. The second purpose of this study was to evaluate the relationship between the percentile category of the STAR test and the TCAP test. The factor variable, identified as the percentile category, included three levels: Urgent Intervention, Intervention, and At/Beyond Benchmark. The dependent variable was the TCAP score. The study included 3rd-grade, 4th-grade, and 5th-grade students during the 2016-2017 school year who had taken the STAR reading and STAR math assessments and had taken the TCAP reading and TCAP math assessment.

Based on the findings of this study, a strong correlational relationship does exist between the STAR and TCAP assessments. Overall, the strong correlation between the STAR and the TCAP were consistent across Math and Reading in 3rd, 4th, and 5th grades. Since the ANOVA was significant, a post hoc multiple comparisons was conducted to evaluate pairwise difference among the means of the three groups. Overall, the At/Beyond Benchmark group was significantly higher than both the Urgent Intervention group and the Intervention group in Math and Reading for 3rd grade, 4th grade, and 5th grade. There was not a significant difference

between the Urgent Intervention group and the Intervention group, the exception was 5th grade math.

DEDICATION

This dissertation is dedicated to God, my family, and my friends. God provided the drive, determination, and means to make this journey possible.

To my husband, Mike, you provided unwavering support, constant encouragement, and served as a pillar of strength throughout this journey. You made many sacrifices and contributions to make this all possible. I thank God for you every day.

To the family and friends God has blessed me with, you have served as an encouragement team that has stepped in when I needed you the most. Your patience, advice, understanding, and encouragement mean more than you will ever know.

Dr. Amanda Cannon and Dr. Anna McCarter, thank you both for always being available to listen when I needed to talk, for answering all the questions I threw your way, and for providing encouragement when I needed it the most. You both serve as an inspiration to me in my career, throughout my educational journey, and in life.

To Dr. Terri Dodge, Peggy Oakes, and Noelle Sutton, your leadership throughout my career has guided me to become a better educator and a better person. I am so thankful God placed you in my life as an administrator but also as a friend.

ACKNOWLEDGEMENTS

I would like to acknowledge and thank:

Dr. William Flora, for giving me the advice, support, assistance, and encouragement needed to see this project through. Your confidence in my abilities served as motivation throughout the doctoral journey.

Dr. Don Good, your statistical knowledge made this possible. Your time and contributions on numerous occasions insured I was moving in the right direction.

Dr. Pamela Scott and Dr. Stephanie Tweed, two members serving on my committee, your feedback and suggestions served as valuable contributions throughout the doctoral process.

Dr. Jason Vance of Loudon County Schools and Dr. Debra Cline of Sevier County Schools, for allowing me the access to principals and scores needed to complete this project.

The principals and school staffers who took personal time to assist me on this journey by collecting the data necessary for this study. Without your efforts, this research study would not have been possible.

The XI State Sorority, for their financial support through multiple member scholarships, for making the pursuit of this endeavor more attainable.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Figure Page

# CHAPTER 1

# INTRODUCTION

Assessments have been part of education for over 100 years and are used for the purpose of assessing student ability and to drive reform (Brewer, Knoeppel, & Lindle, 2015; Linn, 2000; Shepard, 2016). The methods and formats of formalized educational assessments have changed over time. Popham (2008) identified the importance of standardized tests and highlights the misuse of scores.

Based on information provided by Tennessee Department of Education (TNDOE, n.d.a), students are expected to complete high-stakes standardized state assessments on a yearly basis. The state assessment is scheduled near the end of the school year. According to the TNDOE standardized state assessments can be utilized to provide students with academic feedback regarding individual academic strengths and to help educators in evaluating and strengthening instructional practices.

With a single state assessment required at the end of the school year, some classroom teachers utilize benchmark testing as a monitoring method throughout the school year (Renaissance Learning, 2015a). Students complete standardized benchmark tests multiple times throughout the school year. Student results can be compared to national norms by their teacher and student progress can be monitored over time (Renaissance Learning, 2015a). The benchmark test results can also be examined in preparation for the yearly standardized state assessment.

In 2016-2017 the TNDOE has adopted a revised assessment format for measuring student outcomes. In light of this new assessment format, the correlation between a standardized benchmark assessment and the state mandated assessment may be questioned. Teachers and administrators depend on the current benchmark tests to assess student progress as they prepare for the end of year standardized state assessment.

## Statement of the Problem

Marzano (2010) pronounced, “all assessments are imprecise to one degree or another” (p. 13). Sloane and Kelly (2003) argued that no assessment provides data that is above criticism. As teachers are using a computerized benchmark assessments to monitor the progress of student learning over the course of a school year, it is important for teachers and administrators to know if benchmark assessments correlate with high-stakes end of year assessments. For teachers and administrators in Tennessee, it is important to know if differences between Tennessee Comprehensive Assessment Program (TCAP) scores when compared by Standardized Test for the Assessment of Reading (STAR) test scores exist.

This study will determine if a significant correlational relationship exists between a universal screener and benchmark assessment from Renaissance Place, identified as STAR, and the end of year state assessment, identified as the TCAP. More specifically, this study will evaluate student achievement on the TCAP in Math and Reading for grades 3, 4, and 5 when comparing students based on STAR performance level: Urgent Intervention, Intervention, and At/Beyond Benchmark (low, middle, high). Therefore, the purpose of this study is to is to determine if a correlation existed between the STAR assessment and the TCAP assessment in Math and Reading for grade 3, grade 4, and grade 5, and to evaluate the relationship between the percentile category of the STAR test and the TCAP test. The factor variable, identified as the percentile category, included three levels: Urgent Intervention, Intervention, and At/Beyond Benchmark. The dependent variable was the TCAP score.

## Research Questions

The following research questions will be used to guide the study:

$RQ_1$: Is there a significant correlation between the STAR Reading test scores and TCAP in Reading for students in 3rd grade?

$RQ_2$: Is there a significant correlation between the STAR Reading test scores and TCAP in Reading for students in 4th grade?

$RQ_3$: Is there a significant correlation between the STAR Reading test scores and TCAP in Reading for students in 5th grade?

$RQ_4$: Is there a significant correlation between the STAR Math test scores and TCAP in Math for students in 3rd grade?

$RQ_5$: Is there a significant correlation between the STAR Math test scores and TCAP in Math for students in 4th grade?

$RQ_6$: Is there a significant correlation between the STAR Math test scores and TCAP in Math for students in 5th grade?

$RQ_7$: Is there a statistically significant difference in 3rd grade TCAP Reading scores as compared by STAR Reading level (Urgent Intervention, Intervention, At/Beyond Benchmark)?

$RQ_8$: Is there a statistically significant difference in 4th grade TCAP Reading scores as compared by STAR Reading level (Urgent Intervention, Intervention, At/Beyond Benchmark)?

$RQ_9$: Is there a statistically significant difference in 5th grade TCAP Reading scores as compared by STAR Reading level (Urgent Intervention, Intervention, At/Beyond Benchmark)?

$RQ_{10}$: Is there a statistically significant difference in 3rd grade TCAP Math scores as compared by STAR Math level (Urgent Intervention, Intervention, At/Beyond Benchmark)?

$RQ_{11}$: Is there a statistically significant difference in 4th grade TCAP Math scores as compared by STAR Math level (Urgent Intervention, Intervention, At/Beyond Benchmark)?

$RQ_{12}$: Is there a statistically significant difference in 5th grade TCAP Math scores as compared by STAR Math level (Urgent Intervention, Intervention, At/Beyond Benchmark)?

## Significance of the Study

This study will enhance the body of research surrounding the use of benchmark assessments and their relationship to required state assessments. Since multiple forms of student assessment are being used in the Tennessee educational system, determining if a correlational relationship exists between the Renaissance STAR test and the TCAP will be useful for teachers and administrators. The results from this quasi-experimental ex post facto quantitative study could impact the use of Renaissance developed assessments as a tool for making instructional decisions based on accurate data. If assessment results from a for-profit educational testing company are being used within school systems as a screening method, the results should correlate with the state formalized assessment for students. If the results from an outside testing company are not correlating with state assessment results, school administrators may need to obtain a new assessment that can be utilized to make instructional decisions for students.

## Definitions of Terms

In education, there are terms that are used in association with educational practices. Below are the operational definitions for terms important to this study.

1. At/beyond benchmark (as a ranking category on a universal screener or a benchmark assessment) – The highest percentile ranking group on a universal screener or benchmark assessment. The at/beyond group of students score in the $26^{th}$ percentile or higher.
2. Computer adaptive test (CAT) – An assessment designed to adjust the difficulty level of questions based on the responses of the test taker. If a test taker answers a question correctly the following question will increase in difficulty (Renaissance Place, 2017).
3. Intervention (as a ranking category on a universal screener or benchmark assessment) – The middle percentile rank on a universal screener or benchmark assessment. The intervention category includes students scoring between the $10^{th}$ and $25^{th}$ percentile.
4. Multi-gate screening - The strategy of using of multiple assessments to evaluate and correctly identify individual student needs (Levitt et al., 2007).
5. Paper pencil test (PPT) – An assessment that is administered on paper and completed with the use of a pencil.
6. Progress monitoring – a scientifically based model to evaluate the effectiveness of instruction and assess a student's academic performance and/or rate of improvement (National Center for Learning Disabilities, 2006).

7. Raw score – The number of questions answered correctly by a test taker without any numerical computation (Tan & Michel, 2001).
8. Renaissance Place STAR assessment (STAR) – A computer-adaptive assessment, designed to be used as a universal screener and progress monitoring tool (Renaissance Place, 2017).
9. Response to Intervention ($RTI^2$) - A multi-tier approach for early identification and support for students with behavior and/or learning needs (RTI Action Network, 2017).
10. Scaled score – Scores that have been transformed by a mathematical process from raw scores to some form of comparable numerical score. (Tan & Michel, 2001).
11. Tennessee Comprehensive Assessment Program (TCAP) – The standardized state test used in public schools in the state of Tennessee (TN Dept. of Ed., n.d.).
12. Tennessee Educator Acceleration Model (TEAM) - approved teacher evaluation model within the state of Tennessee (TSBOE, 2003).
13. Tennessee Value-Added Assessment System (TVAAS) – A method of measuring how much a student can grow in their academic progress over the course of a school year (TEAM, n.d.c)
14. Tier 1 – All students receive high quality core instruction, which is provided by certified personnel in a classroom setting (RTI Action Network, 2017).
15. Tier 2 – Students not making expected progress based on core classroom instruction alone. In addition to core instruction, students within this group receive additional support, instruction, or intervention to increase their performance in the needed area (RTI Action Network, 2017).

16. Tier 3 – In addition to core instruction the students within this group receive intense individualized intervention to increase students' understanding of deficit skill(s) (RTI Action Network, 2017).
17. Universal Screener – Systematic testing of all students to help in identifying students that may be at risk of falling behind or that have fallen behind grade level expectations (RTI Action Network, 2017).
18. Urgent intervention (as a ranking category on a universal screener or benchmark assessment) – The lowest percentile rank on a universal screener or benchmark assessment. Urgent intervention are students scoring between the 0 and 9$^{th}$ percentile.

## Limitations and Delimitations

The following are limitations of the study. Collection of data will be limited to schools using STAR and TCAP testing platforms. Schools not taking both the STAR and TCAP during the 2016-2017 school year will be excluded from the study. The researcher assumes students receive appropriate tier intervention based on STAR testing data and additional student measures. Researcher assumes accuracy of data sets collected from administrators and that the accuracy of these data sets is further based on principals' accurately reporting assessment data. The schools utilized in this study were contingent on the willingness of administrators to fully participate.

The delimitations of the study include: all schools taking part in the study will come from two participating school districts in the southeast region of Tennessee. The study will only include students in the 3$^{rd}$, 4$^{th}$, and 5$^{th}$ grade. Students not receiving test scores from both testing platforms will be excluded from the research study, and data for the study is limited to the 2016-2017 year.

## Overview of Study

The purpose of this research was to correlational relationship between STAR test scores and TCAP test scores in 3$^{rd}$ grade, 4$^{th}$ grade, and 5$^{th}$ grade in Math and Reading and to evaluate the difference between TCAP scores when compared by STAR Math or Reading level (Urgent Intervention, Intervention, At/Beyond Benchmark). The sampling methodology is convenience sampling. This method was utilized due to the availability of schools using both testing platforms, a willingness to share students testing scores, and geographical location. All of the included testing scores are from students attending participating rural schools from two districts in Southeast Tennessee during the 2016-2017 school year. Data was collected by administrators or testing coordinators at the schools and then provided to the researcher for use in the study. Each data set contained a student's scaled score from the STAR assessment and their scaled score from the TCAP assessment.

# CHAPTER 2

# REVIEW OF LITERATURE

## History of Educational Assessments

According to Haladyna, Haas, and Allison (1998), the first achievement test dates back to the late 1800's. The earliest form of standardized testing was established when fundamental educational philosophies evolved from education of the privileged to education for all (Paulina, 2017). The earliest assessment instruments were used as a way to evaluate individual achievement capabilities. Measuring IQ was the focus of early assessment; the results of an IQ assessment could help identify if a child was functioning normally or if a child was in need of special support (Shepard, 2016).

The NEA article (2017) reported that by 1875 teachers and administrators were creating their own exams to test student preparedness for college. It was not until the College Entrance Examination Board was established in 1900 that a single examination was used nationwide. By 1918, researchers had developed over 100 standardized tests to measure achievement in elementary and secondary subjects (United States, 1992). In 1926, the first Scholastic Aptitude Test (SAT) was administered to college applicants as a method for determining those who were qualified for college level work (United States, 1992). During the early 1900's multiple choice testing was well-established in schools, and by 1935 high speed computing would be applied to standardized tests (U.S. Congress, 1992). Higgins (2009) credits Chauncey, Conant, Johnson and others with creating an educational system that has utilized standardized testing as a principal component of assessment for over seventy years.

In 1965, the Elementary and Secondary Education Act (ESEA) contained specific requirements for the evaluation and accountability of programs for organizations were to receiving funds as Title I programs. The requirement for evaluation and accountability in order to receive funding resulted in an expansion in the use of standardized assessments. Due to the new evaluation requirements, some teachers and administrators started testing students twice a year, in the fall term and again in the spring term (Linn, 2000).

On January 8, 2002, the No Child Left Behind Act (NCLB) was signed into legislation. Lee and Reeves (2012) suggested NCLB legislation established a framework from which educators were charged with improving student achievement in reading and mathematics. In addition to improving student proficiency, NCLB legislation also established requirements to close the academic achievement gap and eliminate the disproportionate academic performance of sub-groups, identified in the following categories: racial/ethnic, socioeconomic, home-language, and special education groups (Dworkin, 2005; Lauen & Gaddis, 2012; Lee & Reeves, 2012). While NCLB legislation delineated teacher expectations for improving scores and closing the achievement gap, the legislation had some unintended consequences. As a direct result of higher expectations for student achievement teachers began to focus instruction on only material that would be covered in end of course or end of year assessments (Dworkin, 2005; Lauen & Gaddis, 2012).

According to information from the National Center for Learning Disabilities (2006), the 1975 Education for All Handicapped Children Act was updated in 2004 as the Individuals with Disabilities Education Act (IDEA). Response to Intervention (RTI) was a key component of the 2004 IDEA reauthorization (Thorius & Maxcy, 2015). The 2004 reauthorization of the IDEA had five new components including more inclusive placements for special education students, an

RTI prevention model, the creation of highly-qualified standards for special education teachers, a focus on assistive and instructional technology, and a universal design for learning (Faieta, 2017).

On December 10, 2015, President Obama signed the Every Student Succeeds Act (ESSA) into law (USDOE, n.d.). The ESSA is the latest evolution of the NCLB Act. The USDOE (n.d.) identify several highlights related to the ESSA, including a requirement for all American students to be taught academic standards that will prepare them for college and careers. The ESSA established broader options for student assessment; these options allowed for student assessments to be conducted throughout the academic year or conducted at the end of the academic year (Collier, 2017).

The use of educational assessment has evolved from assessing student abilities for identifying student placements, to the utilization of student assessments as a measure of teacher evaluation, and accountability for student growth (United States, 1992). Current issues associated with educational assessment range from test format, effect on student motivation and morale, the relationship between standards and assessment, and the difference between "assessment *of* learning and assessment *for* learning" (Sloane & Kelly, 2003, p. 12).

## Purpose of Educational Assessments

Assessments and formalized testing have served as a driving force for educational reform for many years (Linn, 2000; Shepard, 2016). The earliest documented formalized educational assessment was given to individuals in order to measure individual achievement (Brewer et al., 2015). Linn (2000) detailed four reasons assessment serves as a foundation for educational reform: assessments are inexpensive, testing can be externally mandated, testing changes can be implemented quickly, and test results are visible.

Since formalized assessments in education are a common practice, it is important to understand fundamental assumptions for why such assessments are utilized. According to Haladyna et al. (1998), formalized testing was put into place for three interrelating purposes. In the 1800s when the United States introduced education for all, standardized testing was used to provide an effective measure for the education received. Standardized testing has also been used to assess educational progress. Finally, Haladyna et al. perceive the third function of standardized testing is to exclude some students from further opportunities related to their education, instead of identifying potential opportunities where intervention might benefit the individual. "Test scores often become the bases for making decisions about retention, promotion, kindergarten entrance, ability grouping, and special education placements," (Haladyna et al., 1998, p. 265).

## Formats of Educational Assessments

### Summative Assessment

Summative assessments are defined as "tests whose purpose is to make a final success/failure decision about a relatively unmodifiable set of instructional activities" (Popham, 2008, p. 9). Summative assessments can include but are not limited to: end of year state assessments, benchmark testing, end of unit or end of chapter tests, end of semester tests, end of course tests, or a college entrance exam such as the ACT or SAT. Summative assessments are used by teachers to evaluate student achievement or student growth after a period of time within a course (Dixson & Worrell, 2016; Marzano, 2010; Nitko & Brookhart, 2007).

Summative assessments can provide important information regarding the overall significance of student achievement within educational programs (Bennett, 2011). Dixson and Worrell (2016) identify three main features of a quality summative assessment: accurately describe student achievement and student growth as part of an accountability system for a teacher, school, district, and state, provide a valid, reliable, and fair measure of progress toward knowledge and skills necessary to be college or career ready, and to take advantage of computer adaptive testing for effective and efficient measurement.

As documented by the TNDOE (n.d.a), the end of year TCAP servers four main purposes. First, the assessment will provide feedback related to academic progress and compare individual results to grade level expectations and peers across the district and the state. The feedback from the summative assessment will build confidence and transparency about individual student preparedness for college or the workforce. Next, the assessment can be used by teachers to reflect on practices and strengthen instructional practices. Finally the assessment results can be used by state and district leaders to drive decisions related to the allocation of resources (n.d.a).

Stiggins (2002), argues that the ability for a student to achieve suffers as a direct result of standardized state testing. The end of year standardized state assessments do not provide teachers with information useful for making instructional decisions to improve current student achievement (Sloan & Kelly, 2003; Stiggins, 2002). Stiggins (2002) continues by criticizing the amount of resources allocated to standardized assessments, and contends the money could be better used to assist teachers and students during the learning process.

## Formative Assessments

Formative assessment of student learning occurs while the student is in the process of learning (Nitko & Brookhart, 2007). Formative assessments are used to evaluate student learning during the lesson which allows teachers to make instructional adjustments to enhance student learning (Marzano, 2010; Nitko & Brookhart, 2007; Popham, 2008). Formative assessments can include: observations, completion of a graphic organizer, thumbs up/down, exit ticket, using individual white boards, using sign language for multiple choice, and partner talk.

The central purpose of formative assessment is instructional improvement, and formative assessment data can be utilized by the teacher to determine whether to progress with a lesson or reteach a portion of the material for clarification (Bennett, 2011). Through the use of formative assessments, teachers and students contribute to a continuous flow of information related to student achievement (Stiggins, 2002). To maximize the benefits of formative assessments, Stiggins (2002) lists four key 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.

With strong evidence to support their findings, Black and Wiliam (2003), assert the improvement of quality formative assessment will also raise the standards of academic achievement. A high quality formative assessment will serve as a tool to increase student mindfulness (Black & Wiliam, 2003). Maximum results from formative assessments can be achieved when students and teachers work together to assess themselves, provide feedback to modify teaching and learning, and teachers adapt teaching practices to meet all student needs (Black & Wiliam, 2010).

The STAR assessment is designed as a formative assessment which can be used to provide teachers with important information related to individual student performance (Renaissance Place, 2015a). The STAR assessment is a computer adaptive test and can be completed by a student, on average, in about 20-30 minutes. Immediately teachers can assess individual strengths and weaknesses of a student and create an individualized academic plan to advance student skills and abilities (Renaissance Place, 2015a).

### Criterion-Referenced Test

According to Renaissance EdWords (2017a), “a criterion-referenced test is designed to measure a student's academic performance against some standard or criteria” (para. 1). On a criterion-referenced test (CRT) the score is calculated based on student ability to demonstrate understanding of a predetermined set of criteria or educational learning standards (McPherson, 2016). Educators use CRTs to establish what knowledge and skills students have learned compared to what student are expected to learn (Bond, 1996). Due to the standards based scoring technique on a criterion referenced test, Great Schools Partnership (2014) alludes to the possibility every student could fail if expectations are not met or every student could receive the highest obtainable score.

Per Hallam, Lyons, Pretti-Frontczak, and Grisham-Brown (2014), student outcomes on a criterion referenced test can expose individual development in relation to content a student is expected to learn. A CRT can be used as a tool to identify a student in need of additional educational services, special instructional needs, or ongoing observation for instructional variations (Hallam et al., 2014). Sloane and Kelly (2003) identify potential concerns related to results of a criterion referenced assessment: powerful and complex content ideas are difficult to associate with a single learning standard and teachers teaching to the test or teaching students testing structure can inflate test scores even though students may not have an understanding of the content.

As described by TNDOE (2015), the TCAP Achievement Test is a CRT that is updated yearly. The test is designed to measure student performance against content standards set forth by the state of Tennessee. Every item on the TCAP assessment is "directly linked to a performance indicator" (TNDOE, 2015, p.3) Performance indicators are developed in an effort to provide clear and consistent learning goals for each subject and grade level (Common Core Standards, 2017). Results from a CRT will indicate which standards a student has already mastered (EdWords, 2017a).

**Norm-Referenced Testing**

Norm-referenced tests (NRT) are designed to compare a test takers performance to other test takers of the same age (Hallam et al., 2014). NRTs are created to draw attention to allow distinction between the performances of different groups of test takers (Bond, 1996). NRTs can be used to rank or order students based on achievement performance. With NRTs the standard for achievement is set by the test group rather than by a state educational department or a group of experts (McPherson, 2016). Student performance can be ranked low achieving or high achieving based on the performance of the group (McPherson, 2016; Great Schools Partnership, 2014).

Educators can use NRTs to gauge where a student is performing developmentally in relation to same aged students with a classroom or across the nation (EdWords, 2017b, Hallam et al., 2014). Student performance on a nationally based NRT can provide parents with information indicating how their child performed in relation to other students nationally (Popham, 2001). While the results of a NRT may not directly impact classroom instruction, it is beneficial for educators to see at what developmental level students are performing (EdWords, 2017b). Educators can use the results of a NRT to ability group students for instructional purposes, to group students to receive remedial or supplemental instruction and to group students for gifted or enrichment programs (Bond, 1996).

According to Hallam et al. (2014) and Popham (2001), NRTs can be used when determining how to distribute funding resources based on relative strengths and weaknesses of a group of students. As identified by the Great Schools Partnership (2014), norm-referenced tests are considered an objective assessment tool that can lessen the likelihood of bias or favoritism when making education decisions. Potential misuse of norm-referenced tests scores include: basing important educational decisions such as promotion or retention on test results and lowered academic expectations from a teacher (Great Schools Partnership, 2014).

**Universal Screener**

A universal screener is an assessment that is administered to all students on their grade level. The universal screening assessments are typically brief (Center on Response to Intervention, 2015), and as indicated by the RTI Action Network (2017), universal screeners are administered three times per year. Vanderheyden (2013) described a universal screener as a forecast of a child's performance on the end of year state assessment. While a screener can be used as an identifier for students in need of extra support or instruction, using a multi-gate screening strategy is crucial, since a single assessment is not always accurate as a sole identifier (Eklund et al., 2009; Sloane & Kelly, 2008). A multi-gate screening strategy is the use of multiple assessments to correctly identify student needs (Levitt et al., 2007). In agreement with Eklund et al. (2009), the Center on Response to Intervention (2015) recognizes the importance of additional testing and progress monitoring as added evidence for student performance and potential student needs.

**Progress Monitoring**

By continuing the process of student monitoring and/or additional testing following the universal screener, specific student needs can be identified and false negatives can be minimized (Salinger, 2016). A false negative is a student falsely identified as at risk or in need of additional assistance based on the results of a single assessment (Levitt et al., 2007). While Eklund et al. (2009) identify the potential risk of over identification of students when using a universal screener, their research found evidence that a universal screening assessment may help identify students earlier than the process of teacher referral. The universal screener is to be used in conjunction with other forms of student assessment to develop the most accurate assessment of a student's ability (Center on Response to Intervention, 2015; Eklund et al., 2009; Salinger, 2016, Vanderheyden, 2013).

The National Center for Learning Disabilities (2006) defines progress monitoring as, "A scientifically based practice used to assess students' academic performance and evaluate the effectiveness of instruction. Progress monitoring can be implemented with individual students or an entire class" (p. 69). Progress monitoring is the frequent evaluation of instructional effectiveness in a student's progress toward a learning goal (Deno et al., 2009). A student's performance level can be identified by a universal screener and performance goals can be established based on the expected advances in student performance (Safer & Fleischman, 2005).

Frequent evaluation of student performance will allow teachers to measure the effectiveness of instructional and interventional practices to maximize student response toward set goals (Center on Response to Intervention, 2015; Deno et al., 2009; Gillam & Justice, 2010; Safer & Fleischman, 2005.)

**Computer Adaptive Test**

Computerized testing has been associated with student assessment for over 50 years (Jiao & Lissitz, 2012). The computer adaptive test (CAT) is designed to effectively and efficiently select appropriate questions for an examinee (Weiss, 2004). The CAT contains a substantial bank of questions; during an assessment each question is chosen from the bank of questions based on previous questions and answers (Jiao & Lissitz, 2012; Weiss, 2004). Some CATs use an item response theory (IRT) model for question selection. In 2012, the National Council on Measurement in Education (NCME) defined IRT as, "A theory of testing based on the relationship between individuals' performances on a test item and the test takers' levels of performance on an overall measure of the ability that item was designed to measure" (para. 77). When using IRT, the questions answered correctly and the difficulty of the questions answered are taken into consideration when calculating a score (NCME, 2012).

Computer adaptive tests (CAT) have multiple advantages in the realm of educational assessment. Every student receives an individualized test that is unique to their performance (Van Horn, 2003), and the administration of the educational assessment is more cost effective when using a computer than the traditional test booklet (Jiao & Lissitz, 2012; Pearson, 2009; Pomplun, Frey, & Becker, 2002). Computerized testing allows for immediate testing and review of results (Jiao & Lissitz, 2012; Pearson, 2009; Pomplun et al., 2002; Van Horn, 2003). Due to the large number of test items within a test bank, a CAT can be given to students on multiple occasions and a student's progress can be examined (Van Horn, 2003). The *STAR Math Technical Manual*, Renaissance Learning (2015a) claims the reliability of testing is improved because the difficulty of the assessment matches individual performance levels. In addition, Renaissance Learning (2015a) asserts the time it takes to complete an assessment is reduced because the adaptive nature of the CAT prevents student exposure to material that is either too difficult or too easy for their ability level.

### Paper-Pencil Test

According to the Great Schools Partnership (2014), a paper-pencil test (PPT) is considered the traditional administration method for standardized assessments. Administration of a paper-pencil test typically requires a test booklet, answer sheet, pencil, and scratch paper for every individual completing the assessment (Zucker, 2004). Once test takers complete a paper based assessment, all materials are collected, organized, and return shipped to the test administration company (Zucker, 2004). As reported by the TNDOE (n.d.a) months will pass before results from a paper-pencil assessment are available.

## Comparing Computer Based Assessments to Paper-Pencil Assessments

**C**omputerized testing has increased over the last 30 years despite the expense associated with updating equipment (Boo & Vispoel, 2012). Shift in test administration from PPT to computerized testing include immediate scores or test results, more flexibility in administration of the assessment, increase in security, reduction or elimination of testing booklets, and the use of multimedia within an assessment (Boo & Vispoel, 2012; Seo & De Jong, 2015; Wang et al., 2008). As computerized assessment become more common, additional features are becoming available for test takers such as computerized tests that can be administered within a system network or even with off-line capabilities (Wang et al., 2008).

Computer based assessments offer several advantages over paper based assessments, but some conflicts are looming in analysis of computerized testing results when compared to paper based testing results. Bennett et al. (2008) and Clariana and Wallace (2002) found a statically significant difference in test results when comparing assessments completed on a computer and assessments completed on paper. Yet Boo and Vispoel (2012), Vispoel et al. (2001) and Wang et al. (2008), were unable to find a statistically significance difference between results of the assessments completed on a computer and the assessments completed on paper.

Other significant differences exist when comparing computer based assessment and paper based assessments. Vispoel et al. (2001) recorded participants taking more time to complete the computer based assessment than the PPT. Despite the extra time taken to complete the computer based assessment, participants considered the computer based assessment less fatiguing than the PPT (Vispoel et al., 2001). When asked if they preferred a computer based test or a paper based test, participants overwhelmingly selected the computer based test (Boo & Vispoel, 2012; Seo & De Jong, 2015; Vispoel et al., 2001).

## Types of Questions on Educational Assessments

According to Popham (1999), in order for test results to show a spread in scores, test designers want to ask questions that everyone will not answer correctly. Consequently standardized assessments are constructed of mostly mid-level difficulty questions (Popham, 1999). In their study Caldwell and Pate (2013) found simple variations in standardized question format increased item difficulty significantly. Participant performance within the study reflected a significant difference in outcomes based on standardization of question format. Results reflected 71% of participants answering the standard scale item correctly but only 47% participants answering the nonstandard scale item correctly (Caldwell & Pate, 2013).

Van de Watering, Gijbels, Dochy, and Van der Rijt (2008) and Birenbaum and Feldman (1998) found students, specifically males prefer multiple choice style questions over more complex and authentic assessment questions like constructed response items. Birenbaum and Feldman noted a more positive attitude exhibited by students toward multiple choice questions when compared to free response style questions (as cited by Traub & MacRury, 1990). Students believed the multiple choice questions were easier to study for, easier to complete, and in turn easier to earn a higher achievement score (Birenbaum & Feldman, 1998). In her research, DeMars (2010), found boys scored higher on multiple choice items, but girls scored higher on constructed response items. McAllister and Guidice (2012) reasoned the value of a test is determined by the careful construction of test questions, not the format of the test questions.

Multiple-choice questions have been the dominate format of assessments within the United States and around the world (Popham, 2008). A question or statement is provided on an assessment and the test taker is to select the correct response from a list of choices, usually four (Marzano, 2010; Nitko & Brookhart, 2007; Popham, 2008). Advantages of multiple choice questions include the ability to assess a wide range of skills in a single assessment and the ease of grading (McAllister & Guidice, 2012). Nitko and Brookhart (2007) indicated the importance of carefully created distractors within a multiple choice question, by asserting the incorrect choices can be used to identify challenges a student may be facing. Popham (2008) argued a weakness of the multiple-choice question format suggests students are not required to generate a correct answer, only recognize the correct answer in a list of choices. Nitko and Brookhart (2007) noted the probability a student would correctly guess multiple choice items is lower than a true false format or a poorly constructed matching exercise.

Matching items are created when two parallel lists of corresponding information are presented (Popham, 2008). Items from one list are connected to items from the second list. Matching can be created with connected words, a word and definition, a word and a symbol or abbreviation, or words and numbers. One disadvantage of matching is the likelihood of low level memorization of items within the list for the exercise of matching (Nitko & Brookhart, 2007; Popham, 2008). Popham (2008) suggests the use of extra responses in matching to limit the likelihood of a student using the process of elimination to solve for unknown premises. The TNDOE (n.d.b) exhibits another form of matching with the matching table. A matching table requires the test taker to match information from the rows within the table to the correct response located in the column of the table (TNDOE, n.d.b). Nitko and Brookhart (2007) suggested a matching table or matrix is useful for a teacher to assess student ability to organize closely related facts or ideas.

Alternative choice (Marzano, 2010), true/false questions, and binary choice (Popham, 2008) are very similar because they are all defined as questions with two possible responses. Teachers often use true/false questions because the questions are easy to write, the questions can be scored easily, and the questions can “cover a wide range of content within a relatively short period” (Nitko & Brookhart, 2007, p. 139). One drawback to alternative choice, true/false, or binary choice is the 50/50 chance of getting a correct answer, even if the test taker has no knowledge over the material being tested (Nitko & Brookhart, 2007; Popham, 2008). This question format has been referred to as “one of the most unreliable forms of assessment” (McAllister & Guidice, 2012, p. 195).

Multiple-response (multiple-select or selected response) questions are formatted similar to multiple-choice, however within the list of choices more than one response is correct (Marzano, 2010). A limitation related to the multiple-response question format falls within the selection of the items responses. As documented by the TNDOE (n.d.h), “sometimes the number of correct responses will be indicated (e.g., “choose the two correct answers”), but sometimes the number of correct responses will not be indicated (e.g., “select all of the correct answers”)” (p. 10).

Fill in the blank questions or short answer questions are completed by contributing a word, phrase, or sentence to complete a statement correctly (Marzano, 2010; Nitko & Brookhart, 2007; Popham, 2008). The assessment taker is required to produce the answer for the question. Nitko and Brookhart (2007) recommend the use of short answer test questions in order to assess student acquisition of knowledge. A student is required to produce the correct answer for a short answer question. The main weakness within this type of question is often related to the length of the response. Since students are required to construct the answers, variations of similar answers may or may not be acceptable when grading (Popham, 2008). Student responses can be difficult to score and variations in scoring can cause an assessment to be less reliable as a measure of student knowledge.

## Reporting Scores on Educational Assessments

One of the most commonly used scoring methods is a percentile or percentile rank. According to Popham (2008), “percentiles are used most frequently in describing standardized test scores because percentiles are readily understandable to most people” (p. 286). A percentile score compares student performance against other students within a norm group, and the percentile score is reflective of the percentage of students outscored by a student’s performance (Popham, 2008). A percentile score of 63% signifies student performance is better than 63% of the other students within the age/norm reference group (Nitko & Brookhart, 2007; Popham, 2008), and norm-referenced tests most often report scores using percentile (Edwords, 2017b).

Instructional reading level (IRL) is the grade level at which a student is at least 80% proficient at recognizing and comprehending reading materials (Renaissance Place, 2017). The IRL is presented using a decimal number. An IRL of 5.6 would mean the child is 80% proficient at reading materials and language appropriate for a 5$^{th}$ grader in the sixth month of school. Renaissance Place formulates student IRL score based on their performance on the STAR Reading assessment (2017).

Renaissance Place (2017) also uses a zone of proximal development (ZPD) score. The ZPD is defined in a range such as 3.7-5.8. The range represented by the ZPD is the area from which a student should be selecting reading materials for optimal growth within reading. The material within the approximate ZPD is ability level appropriate, yet provides a suitable challenge for a student (Renaissance Place, 2017). The ZPD score can be understood as years for the whole number and months for the decimal number. For example, a ZPD of 3.7-5.8 would indicate the students optimal reading level is material rated for a third grader in the seventh month through a fifth grader in the eighth month (Renaissance Place, 2017).

A raw score is another way of scoring student performance on a standardized assessment. A raw score is the total number of questions a test taker answered correctly (Popham, 2008; Tan

& Michel, 2001). If an assessment has a total of 25 questions and a student answers 23 out of the 25 correctly, the raw score would be 23.

A scaled score is a numerical score produced when the raw score is converted mathematically into a new statistically comparable number (Popham, 2008; Tan & Michel, 2001). The conversion method a scaled score is based on an IRT (Popham, 2008). Within the IRT, each test item is individually weighted based on item difficulty and other technical properties. The example presented by Popham (2008) reads, “A student who gets a raw score of 35 correct out of 50 items, for example, might end up with a converted scaled score of 620” (p. 295).

## Response to Intervention (RTI)

Response to Intervention was created to promote the success of students with a disability within the general educational classroom by offering a tiered approach to student interventions (Faieta, 2017). Little (2012) identified RTI as “a systematic data-based method for identifying, defining, and resolving students’ academic or behavioral difficulties” (p. 69). The RTI system is structured as a multitiered model for instructional practices. The layers of the interventional model start in general education, and each increasing level within the model provides more intensified instruction (Fletcher & Vaughn, 2009). Wixson and Valencia (2011) claimed the intent of RTI is to differentiate instruction through more individualized instruction to meet specific student needs, and as a result of the individualized instruction, student learning can be accelerated. RTI has two main goals: “deliver evidence-based instruction and interventions to improve student learning and to collect information regarding students’ responses to those interventions” (Little, 2012, p. 71).

The RTI Action Network (2017) identified the three tiered levels of instruction in the RTI model as Tier 1, Tier 2, and Tier 3. One key component of the RTI system is the need for a universal screening measure for all students (Fletcher & Vaughn, 2009). An assessment used as a screener gathers data before instruction to help identify any student that may be performing below expected age or grade level (Wixson & Valencia, 2011). The required norm-referenced or criterion referenced screener can over identify the students considered at risk. These students are more closely monitored or even provided with differentiated instruction to ensure their success (Fletcher & Vaughan, 2009).

In the RTI system, Tier 1 consists of high quality research-based core instruction inside the general education class (Crepeau-Hobson & Biance, 2012; Denton, 2012; Fletcher & Vaughn, 2009; RTI Action Network, 2017). Students in Tier 1 are monitored closely and participate in systematic screenings to identify any student that is at-risk of falling behind (Denton, 2012). Students identified as at-risk should receive additional support and supplemental materials in the regular classroom during the school day (RTI Action Network, 2017). If one intervention is not working, the intervention can be changed while student progress is closely monitored (Hale, 2008). The length of these interventions can vary, but should not go beyond eight weeks. If student progress is not adequate, the student may be moved to Tier 2 (RTI Action Network, 2017).

Tier 2 is small group instruction within the regular educational classroom in the deficit areas of struggling students (Crepeau-Hobson & Biance, 2012). Fletcher and Vaughan (2009) identified a more specific regimen of groups no larger than three to five students and time frame of twenty to forty minutes daily. Selecting the appropriate intervention plan is critical. Hale (2008) suggested the use of a specific intervention that has shown success for other students with similar struggles. Students continuing to show little or no progress at this level should be considered for more intense interventions as part of the RTI's Tier 3 (RTI Action Network, 2017).

A Tier 3 intervention is more intensive and more focused on student deficit areas than Tier 2. Increased interventional instruction can go up to 45-60 minutes with a more specialized teacher providing interventional instruction, typically outside of the regular classroom (Fletcher & Vaughan, 2009). The tiered interventions do not replace the core classroom instruction. Instead the intervention is provided to students as a supplemental addition to core instruction (Denton, 2012). This increases student opportunity for exposure and practice.

According to the information provided by the RTI Action Network (2017), "Students who do not achieve the desired level of progress in response to these targeted interventions are then referred for a comprehensive evaluation and considered for eligibility for special education services." The movement through the tiers helps ensure each child has adequate instruction and lack of instruction is not the cause of the deficit in skill (Fletcher & Vaughan, 2009). Following the comprehensive evaluation and the review of multiple data sources, a school team can meet and make a decision about a child's eligibility for special education services (Hale, 2008).

## Uses of STAR as an Educational Assessment

According to Renaissance Place (2017), the STAR Reading assessment has three main purposes:

> First, it provides educators with quick and accurate estimates of reading comprehension using students' instructional reading levels. Second, it assesses reading achievement

relative to national norms. Third, it provides the means for tracking growth in a consistent manner longitudinally for all students (para. 1).

The STAR Reading test is a CAT and typically takes less than 30 minutes for students to complete (Renaissance Learning, 2015b). Based on the *STAR Reading Technical Manual,* the CAT design "minimizes frustration and provides more accurate scores for both high-performing and low-performing students" (p. 1). The STAR Reading assessment consists of 34 computer adaptive multiple choice questions. The test does not have a time limit, but the individual items on the assessment have a maximum amount of time allowed per item (Renaissance Learning, 2015b). The last five questions on the STAR Reading assessment are longer passages and more time is permitted for these questions (Renaissance Learning, 2015b).

The Renaissance Learning (2015a) *STAR Math Technical Manual* identifies the structure of the STAR Math assessment as a CAT consisting of multiple choice questions. The STAR Math test is a 34-item assessment which draws from a bank of over 4,000 questions measuring more than 550 skills (Renaissance Learning, 2015a, p. 5). Per the manual, each item was created so that a maximum of one minute would be needed to solve, if the student knew how to do the required math to find the solution. The STAR Math assessment allows a maximum of three minutes per item and a warning is provided to students when only fifteen seconds remain for an item (Renaissance Learning, 2015a).

The STAR Math and the STAR Reading assessments are computer adaptive tests utilizing an IRT for question selection (Renaissance Learning 2015a, 2015b). The IRT design of the CAT "minimizes frustration and provides more accurate scores for both high-performing and low-performing students" (Renaissance Learning, 2015b, p. 1). Upon completion of the STAR Reading assessment, a teacher can immediately access norm-referenced scores such as grade equivalence, scaled score, and percentile rank and criterion-referenced scores for a student such as instructional reading level and zone of proximal development (2015b). The STAR Math assessment provides teachers with the following norm-referenced scores: grade equivalence, scaled score, and percentile rank.

Schools utilize the STAR Math and STAR Reading assessment as a screener to test all students two to four times per year, as outlined by Renaissance Learning 2015a). At the beginning of the school year the STAR assessment is used as a screener to detect students not performing at/on grade level. Students falling between the 0-10th percentile rank are categorized as Urgent Intervention, students falling between the 11th-25th percentile rank are categorized as Intervention, and students performing at the 26th 100th percentile are categorized as At/Beyond Benchmark for RTI.

In addition to using the STAR program as a universal screener and an assessment tool, teachers can also access valuable and specific information regarding suggested skills for instruction for each individual student (Renaissance Learning, 2015b). Upon completion of a STAR Math or STAR Reading assessment an Instructional Planning Report is available for each individual student (See Appendix C STAR Math or Appendix D STAR Reading). Each student will have a detailed list of suggested skills. Based on the information provided by Renaissance Learning (2015a), the list of suggested skills is individualized for each student based on their most recent performance on a STAR assessment and the progression of learning within a subject area. Teachers can create an individualized instructional plan for each student and set specific individualized learning goals to meet the needs of each student (Renaissance Learning, 2015a; 2015b).

Within the STAR program teachers have the ability to set goals for individual students, a small group, or a whole class. Actions teachers can take to maximize the effectiveness of assessments includes: communicating and emphasizing student learning goals and providing practical feedback related to the goals (Harlen & Deakin-Crick, 2002). Renaissance Learning (2015a), claims “Goal setting is an almost ubiquitous practice in education” (p. 119). McTighe and O'Connor (2005) found the “most effective learners set personal learning goals, employ proven strategies, and self-assess their work” (p. 16). It is the responsibility of teachers to promote, model, and expect these successful practices within the classroom for all students (McTighe & O'Connor, 2005).

**Tennessee Comprehensive Assessment Program (TCAP)**

In Tennessee, the TCAP has been the states formal assessment choice since 1988 (TNDOE, n.d.c). Based on the information provided by the Tennessee Department of Education website, the state assessment serves six main objectives:

1. Provide feedback about students’ academic progress and how it aligns with grade-level expectations
2. Give parents and teachers a big-picture perspective about how a student is progressing compared to peers across the district and state, including student strengths and growth opportunities
3. Build confidence and transparency about student readiness for postsecondary and the workforce among Tennessee colleges, universities, and employers
4. Help educators strengthen instruction and reflect on their practice
5. Hold us accountable to serving all students fairly

6. Highlight schools where students are excelling, so we can learn from those who are doing well (n.d.e)

The TCAP is mandatory in connection with school and district accountability as required by the Every Student Succeeds Act (ESSA) of 2015 (TNDOE., n.d.c). The design of the TCAP test, TNReady, assesses true understanding of state standards in each subject and grade level, it is a criterion referenced assessment (TNDOE, 2015).

According to the TNDOE (n.d.b), all questions created for the state assessment go through a dynamic three step evaluation process. First, the department, teachers, and the test vending company cooperatively create test questions based on the academic standards. Next, the created questions are examined by Tennessee teachers and the department. During this phase, the questions can be accepted, revised, or even rejected. Finally, a test is compiled by the Educational Department and test vending companies. All questions are field tested and reviewed for statistical validity before being added to an operational state assessment (TNDOE, n.d.b).

Based on the 2017 Tennessee Every Student Succeeds Act (ESSA) State Plan, "the 2016-17 state assessments will continue to feature multiple types of questions that measure the depth of Tennessee Academic Standards, specifically students' problem solving and critical thinking skills" (p. 33). The TNDOE (n.d.g) indicates, TNReady is an element of TCAP. The TNReady portion of the assessment is a change in assessment design to better measure student understanding of material instead of memorization or test taking abilities. TNReady is an element of the state achievement assessment in select subject areas, as the state transitions to higher academic standards (TNDOE, n.d.d).

Questions featured on the TNReady TCAP assessment are no longer limited to multiple choice questions (TNDOE, n.d.d). The English/Language Arts (ELA) portion of the TCAP includes multiple choice items, multiple select items, and written responses or short answer to support an answer. In the Math section of the TCAP assessment, students encounter fill in the blank questions, short answers items, and questions that will not allow a calculator. The more rigorous question formats that may also be included on the TCAP include: selected response, multiple select, and drag and drop items (TNDOE, 2017a).

## Time Spent on Testing

The Tennessee Department of Education (n.d.f) discloses the amount of time required for end of year state assessments: in 3rd grade English contains four subparts for a total of 216 minutes, Math contains 3 subparts for a total of 115 minutes, Science is one subpart lasting 50 minutes, and Social Studies is one subpart lasting 50 minutes. In 4th grade English contains four subparts for a total of 222 minutes, Math contains 3 subparts for a total of 115 minutes, Science is one subpart lasting 50 minutes, and Social Studies is one subpart lasting 50 minutes. In the 5th grade English contains four subparts for a total of 200 minutes, Math contains 3 subparts for a total of 115 minutes, Science two subparts lasting 95 minutes total, and Social Studies contains two subparts lasting 100 minutes total. The state mandated testing will require a total of 431 minutes for a 3rd grade student, 437 minute for a 4th grade student, and 510 minutes for a 5th grade student (TNDOE, n.d.f).

In addition to state mandated testing, many districts participate in benchmark testing and other district required assessments (Nelson, 2013). Based on Nelson's findings, some school districts were spending roughly 15 hours per year on state and district required assessments. State Collaborative on Reforming Education (SCORE) issues "a report on the state of assessments in Tennessee, informed by extensive feedback from Tennessee's teachers, principals, and district leaders" (p. 3). Based on the SCORE (2015) findings 51% of district leaders, 56% of principals, and 74% of teachers believe too much time is spent on assessments.

Southern (2015) reported, "teachers identified lost instructional time as a result of both district- and state-level assessments as a top challenge faced in their roles" (para. 10). In addition to loosing instructional time due to the multiple assessments, teachers are also faced with the challenge of preparing students for the format of various assessments (Southern, 2015). Nelson (2013) identified test preparation as the administration of practice tests and the teaching of test taking strategies. In preparing for the state mandated assessment, Nelson (2013) reported some school districts spend as much as 80 hours per year on test preparations. As indicated by teacher responses in the report released by SCORE (2015), the number one challenge facing teachers as result of required standardized assessments is the reduction in instructional time.

## Uses of Assessment Data

Assessment for learning and assessment of learning are both essential elements to maximize student achievement (Stiggins, 2002). Guskey (2003) classifies large scale standardized assessments as a useful tool for ranking schools or students for accountability purposes, but argues these large scale assessments are not beneficial to teachers for improving instruction. Three main reasons these large scale assessments are not improving teacher instruction: students complete the assessment at the end of the year after a majority of instructional activities are completed or near completion, scores are not available for months and students have typically moved on to another teacher, and results from the assessment usually lack the level of detail to target specific improvement areas (Guskey, 2003).

According to Stiggins (2002), "Assessments of learning have been the norm throughout the U.S. for decades" (p. 759). Policy makers offer rewards for school producing high scores on standardized test and sanctions for schools that do not produce the desired results. Nelson (2013) noted most large scale standardized tests are summative in nature and are most often used for district, school, or teacher accountability purposes. A concern related to standardized assessments lies in defining the purpose for which assessments are created and administered (Solorzano, 2008). If an assessment is designed to evaluate student achievement, the same assessment may not accurately reflect teacher, school, or district effectiveness (Bond, 1996; Guskey, 2003; Solorzano, 2008; Tienken, 2015). Popham (1999) and Tienken (2015) argue against using summative assessments to judge the quality of educational instruction.

> Employing standardized achievement tests to ascertain educational quality is like measuring temperature with a tablespoon. Tablespoons have a different measurement mission than indicating how hot or cold something is. Standardized achievement tests have a different measurement mission than indicating how good or bad a school is. (Popham, 1999, p. 10)

In one instance, Tienken (2015) found eight conclusions drawn from the results of a single standardized state assessment.

> The results from the state-mandated high school mathematics test in Grade 11 could be used to make determinations about (a) the effectiveness of the high school principal, (b) the effectiveness of the high school math teachers, (c) the quality of the school district's mathematics program, (d) whether a Grade 11 student is college ready, (e) whether that student is career ready, (f) a student's strengths and weaknesses in math, (g) Grade 12

course placements for that student, and (h) whether the student can graduate high school. (p. 156)

Based on teacher survey results, Southern (2015) identifies the top five uses of assessments within the classroom by a teacher: diagnose student skill deficits, guide improve to teacher instruction, set goals with students, to group students, and to predict future student performance. Assessments most useful in improving student learning are the writing assignment, quizzes, tests and other assessments administered by the classroom teacher on a regular basis (Guskey, 2003). Stiggins (2002) thought this assessment for learning process is beneficial to student and teachers. Stiggins (2002) and Risko and Walker-Dalhouse (2010) argued for unprecedented gains in student achievement if teachers would use classroom assessments as a tool to revise instructional practices. Teachers can evaluate classroom assessments, identify areas of strength and weakness, and make instructional decisions in the best interest of the students (Guskey, 2003). The assessments used to evaluate student learning during the lesson which allows teachers to make instructional adjustments are called formative assessments (Marzano, 2010; Nitko & Brookhart, 2007; Popham, 2008).

Formative assessments are those assessments created as instructional tools to promote student learning, and Risko and Walker-Dalhouse (2010) reasoned every classroom assessment should be directly linked to an instructional objective. Wiliam (2017) claimed, the use of assessment during the learning process as opposed to an assessment at the completion of teaching, will have a greater impact on how quickly students learn than almost any other factor. Based on information provided by Tomlinson and Moon (2013), formative assessments help teachers identify patterns in student performance in an effort to enhance teacher instruction. According to Wiliam (2011), the assessment process is necessary in order for teachers to make adjustment to instructional practices. Wiliam (2017) proposed a simple three question process to drive effective use of formative instruction: Where is the learner? Where does the learner need to go? How do we get there?

Upon the completion of a formative assessment, a student needs to be provided with quality feedback, which often includes a plan of action for goal setting and future academic growth (Wiliam, 2017). When a student completes a formative assessment, allowing the student to actively participate in the analyzation of the assessment results will allow the student to take ownership of their learning (Tomlinson & Moon, 2013). The quality of the interaction between the teacher, the student, and the assessment will be the basis of any outcome or change (Black & Wiliam, 1998).

Wiliam and Black (1998) recorded students were able to make large and meaningful gains when teachers used assessment to drive instructional practices. As a guide for the cooperation of teachers and students in the effective use of assessment, Wiliam (2017) suggests:

> "Assessment improves learning when it is used to support five key strategies in learning: Clarifying, sharing and understanding learning intentions and criteria for success. Engineering classroom discussions, activities and tasks that elicit evidence of student achievement. Providing feedback that moves learning forward. Activating students as learning resources for one another. Activating students as owners of their own learning. (p. 5)

## Teacher Evaluations and Assessments

According to the Tennessee State Board of Education (TSBOE) all schools are required to adopt an approved teacher evaluation model (2017). "The primary purpose of annual teacher and school administrator evaluation is to identify and support instruction that will lead to high levels of student achievement” (TSBOE, 2017, p.1). The Tennessee Educator Acceleration Model (TEAM) is one of the approved teacher evaluation models within the state of Tennessee.

The design of the TEAM is 50% quantitative and 50% qualitative. The qualitative evaluation measure of the TEAM is based on teacher observation TEAM (n.d.d). Teacher observations are completed at the school level by TEAM certified evaluators (TSBOE, 2017). The number of observation is determined by licensure status and previous year individual growth or overall evaluation score. All observations are followed with a post-conference meeting with the certified evaluator to provide the teacher with constructive feedback (TEAM, n.d.d).

The quantitative evaluation measure of the TEAM, the remaining 50% of the teacher evaluation measure, contains two forms of student achievement data from standardized testing (TEAM, n.d.b.). The two forms of student achievement measures combine student achievement and student growth (TEAM, n.d.a). Thirty-five percent of the student achievement measure comes from student growth as measured by TVAAS TEAM (n.d.b). The remaining 15% of the student achievement measure comes from a teacher selected form of student achievement as approved by the state (TSBOE, 2017).

“TVAAS allows educators to consider their students’ achievement (their score on the end of year assessment), as well as their growth (the progress students make year to year)” (TEAM, n.d.c, para. 2). Through TVAAS, educators are held accountable for student growth during the current school year only. Whether students start the year high or low achieving, the growth over the course of the school year is the focus of the TVAAS (TEAM, n.d.c).

When calculating a teacher’s TVAAS score, students are not measured individually, instead the whole class is considered a group (TEAM, n.d.c). A determination about the groups starting point is calculated based on the group's performance on previous years standardized achievement tests. The group scores from previous years are then compared to the end of year standardized assessment to get an idea of a teacher's performance. By using the scores of groups of students TVAAS can provide a strong and reliable basis for determining the progress of students within classrooms, schools, and districts (TEAM, n.d.c).

The remaining 15% of the quantitative students achievement measure is teacher selected (TEAM, n.d.a). Information provided by TEAM (n.d.a) states:

> The State Board of Education approved options for teachers and principals for the 15 percent achievement measure component that reflects those measures that showed a relationship to student growth and that could be returned in a timely manner. Teachers should meet with their evaluators early in the school year to choose a 15 percent measure and set clear and rigorous goals. (para. 1)

Within the guidelines provided, teachers can select the remaining 15% of the quantitative students achievement measure by choosing: state assessment growth, state assessment achievement, TVAAS, off the shelf assessment (including STAR) growth, or off the shelf assessment achievement (TEAM, n.d.a).

## Negative Impact of Standardized Assessments

When assessment results are used correctly education can be improved; however the incorrect use of standardized test results will continue to negatively impact the educational experience (Guskey, 2003). From using a standardized assessment as a single measure of graduation requirement (Cizek, 2001), to limiting the content covered within a class (Roderick &

Engel, 2001) the negative impacts of standardized testing are negatively impacting schools across the nation.

Cizek (2001) highlighted the rapid change in testing structure in Washington, Arizona, and Massachusetts. Standardized tests in these states went from low-level recall type questions to complex high-level content questions considered too challenging for the test takers. Also notable, educators are presenting challenging contextual ideas to students too early due to the swift changes to the content standards in preparation for the new assessment (Cizek, 2001).

Popham (1999) argued the design of some standardized test questions favor students from higher socioeconomic status. Student exposure, life experience, and prior knowledge will result in certain test questions easier for select students, while a student with limited exposure may not possess necessary understanding of a questions pieces (Popham, 1999). Roderick and Engel (2001) speculate the pressures related to high-stakes testing in schools can cause students immense anxiety and for some students even cause disengagement from school or academics.

The negative impacts of standardized assessments may also affect educators. Cizek (2001) emphasized the impact on content selection. Often educators are forced to limit the amount of time spent on content not covered by assessment, restrict or omit time spent on life lessons or daily life skills, and reduce the amount of time students spend on each instructional activity (Cizek, 2001). Nichols and Berliner (2008) quoted Campbell's Law, "The more any quantitative social indicator is used for social decision-making, the more subject it will be to corruption pressures and the more apt it will be to distort and corrupt the social processes it is intended to monitor" (p. 672). Nichols and Berliner (2008) cautioned cheating on tests, data manipulation, teaching to the test, and the demoralization of teachers are all direct results of the pressure educators feel to do well on standardized assessments.

Popham (2001) argues against using standardized achievement test for the evaluation of teachers or schools. In order to create deviations in student performance on standardized tests, Popham alleges:

> the developers of such tests sometimes include items apt to be answered correctly by students who either (a) come from advantaged socioeconomic backgrounds or (b) were fortunate enough to inherit above-average academic aptitudes such as verbal or quantitative capacities. Test items based chiefly on these two factors almost always produce the desired spread of student scores because both socioeconomic status and inherited academic aptitudes reflect what children bring to school, not what they learn there. (2001, p. 27)

Using student assessment scores as a form of teacher evaluation may also come with unintended consequences (Knight et al., 2012). “Anyone familiar with schools realizes that the caliber of a given teacher's students can vary dramatically from year to year” (Popham, 2001, p.28). The use of student test scores as a method for evaluating teacher effectiveness neglects to account for several influencing factors: class size, support within the home or community, individual student health needs which may impact attendance, previous teachers or schooling, and summer learning loss which impacts students at different rates (Darling-Hammond, Amrein-Beardsley, Haertel, and Rothstein, 2012). Goldhaber, expresses the potential for prospective teachers to select another profession that does not face the same evaluation scrutiny (2015).

Other potential consequences of using student assessments as an evaluation method for teachers might also negatively impact schools and communities (Knight et al., 2012). Teachers, likely to feel the pressure for their students to perform well on standardized assessments, may alter their teaching methods or limit their additional tasks or responsibilities. According to

Knight et al. (2012), experienced teachers may be less likely to mentor novice teachers, teachers may be less willing to spend their time as a mentor or tutor, and limit the curriculum to only material on the test.

## Chapter Summary

With numerous assessment types and a range of scoring methods for these assessments available, it is critical to understand the differences (Hallam et al., 2014). Content selection and scoring procedures for a CRT are based on a predetermined set of standards or criteria (EdWords, 2017a), but ranking methods for norm-referenced tests are dependent on a peer group performance (EdWords, 2017b). When using CATs, each student receives an individualized assessment (Van Horn, 2003), administering the assessment is more cost effective than a paper-pencil test (Jiao & Lissitz, 2012; Pearson, 2009; Pomplun et al., 2002), and results are available immediately (Jiao & Lissitz, 2012; Pearson, 2009; Pomplun et al., 2002; Van Horn, 2003).

Due to the impact and influence of test scores, it is the responsibility of teachers, administrators, and district educational officials to carefully select and develop appropriate educational assessments (Bond, 1996). Since a single assessment cannot measure everything considered important (Bond, 1996), it is important to evaluation multiple forms of assessments to increase understanding of student ability and needs (Edwards et al., 2008).

# CHAPTER 3

# METHODOLOGY

The primary purpose of this quasi-experimental ex post facto study was to correlate the scaled score from the Renaissance Place STAR assessment to TCAP in Math and Reading for 3rd grade, 4th grade, and 5th grade. The secondary was to evaluate the difference between TCAP scores when compared by STAR Math or Reading level (Urgent Intervention, Intervention, At/Beyond Benchmark). This chapter contains information that can be utilized for research design, population, instrumentation, procedures, and data analysis.

## Research Questions and Null Hypotheses

The following research questions and corresponding null hypotheses were used to guide the study:

$RQ_1$: Is there a significant correlation between the Renaissance STAR Reading test scores and TCAP in Reading for students in 3rd grade?

$H_{01}$: There is not a significant correlation between the Renaissance STAR Reading test scores and TCAP in Reading for students in 3rd grade.

$RQ_2$: Is there a significant correlation between the Renaissance STAR Reading test scores and TCAP in Reading for students in 4th grade?

$H_{02}$: There is not a significant correlation between the Renaissance STAR Reading test scores and TCAP in Reading for students in 4th grade.

$RQ_3$: Is there a significant correlation between the Renaissance STAR Reading test scores and TCAP in Reading for students in 5th grade?

$H_{03}$: There is not a significant correlation between the Renaissance STAR Reading test scores and TCAP in Reading for students in 5th grade.

$RQ_4$: Is there a significant correlation between the Renaissance STAR Math test scores and TCAP in Math for students in 3rd grade?

$H_{04}$: There is not a significant correlation between the Renaissance STAR Math test scores and TCAP in Math for students in 3rd grade.

$RQ_5$: Is there a significant correlation between the Renaissance STAR Math test scores and TCAP in Math for students in 4th grade?

$H_{05}$: There is not a significant correlation between the Renaissance STAR Math test scores and TCAP in Math for students in 4th grade.

$RQ_6$: Is there a significant correlation between the Renaissance STAR Math test scores and TCAP in Math for students in 5th grade?

$H_{06}$: There is not a significant correlation between the Renaissance STAR Math test scores and TCAP in Math for students in 5th grade.

$RQ_7$: Is there a statistically significant difference in 3rd grade TCAP Reading scores as compared by STAR Reading level (Urgent Intervention, Intervention, At/Beyond Benchmark)?

$H_{07}$: There is not a statistically significant difference in 3rd grade TCAP Reading scores as compared by STAR Reading level (Urgent Intervention, Intervention, At/Beyond Benchmark).

$RQ_8$: Is there a statistically significant difference in 4th grade TCAP Reading scores as compared by STAR Reading level (Urgent Intervention, Intervention, At/Beyond Benchmark)?

$H_{08}$: There is not a statistically significant difference in 4th grade TCAP Reading scores as compared by STAR Reading level (Urgent Intervention, Intervention, At/Beyond Benchmark).

$RQ_9$: Is there a statistically significant difference in 5th grade TCAP Reading scores as compared by STAR Reading level (Urgent Intervention, Intervention, At/Beyond Benchmark)?

$H_{09}$: There is not a statistically significant difference in 5th grade TCAP Reading scores as compared by STAR Reading level (Urgent Intervention, Intervention, At/Beyond Benchmark).

$RQ_{10}$: Is there a statistically significant difference in 3rd grade TCAP Math scores as compared by STAR Math level (Urgent Intervention, Intervention, At/Beyond Benchmark)?

$H_{010}$: There is not a statistically significant difference in 3rd grade TCAP Math scores as compared by STAR Math level (Urgent Intervention, Intervention, At/Beyond Benchmark).

$RQ_{11}$: Is there a statistically significant difference in 4th grade TCAP Math scores as compared by STAR Math level (Urgent Intervention, Intervention, At/Beyond Benchmark)?

$H_{011}$: There is not a statistically significant difference in 4th grade TCAP Math scores as compared by STAR Math level (Urgent Intervention, Intervention, At/Beyond Benchmark).

$RQ_{12}$: Is there a statistically significant difference in 5th grade TCAP Math scores as compared by STAR Math level (Urgent Intervention, Intervention, At/Beyond Benchmark)?

$H_{012}$: There is not a statistically significant difference in 5th grade TCAP Math scores as compared by STAR Math level (Urgent Intervention, Intervention, At/Beyond Benchmark).

## Instrumentation

The researcher used existing test scores from two different standardized assessments, STAR and TCAP. The STAR Math and STAR Reading assessments are administered periodically during the school year as a form of benchmark testing or progress monitoring (Renaissance Learning, 2015a). The STAR assessments are computer adaptive and are designed to provide teachers with immediate and individualized feedback for students. According to Renaissance Learning, the STAR Math assessment can be completed in about 20 minutes (2015a), and the STAR Reading assessment can be completed on average in under 30 minutes (2015b). Once a student has completed the assessment, the program will provide immediate and individualized feedback for that student, a small group of students, or all the students in that classroom (Renaissance Place, 2015a). One evaluation score from the STAR assessment is a scaled score. As noted by Renaissance Place (2017), the scaled score is calculated based on the difficulty of the questions and the total number of correct responses within the assessment.

The TCAP is an end of year state assessment used in Tennessee. According to the information provided by Tennessee Department of Education (n.d.d), the design of the state assessment is intended to assess true student understanding. The state assessment “offers parents, students, and teachers with an academic check-up each year to ensure all students are moving forward, on track to be successful in the next step of their academic journey" (p. 5). The times for 2016-2017 TCAP assessment are Reading 195 minutes and Math 115 minutes for 3rd grade, 4th grade, and 5th grade. In 2016-2017 the paper pencil TCAP assessment was created using a combination of fill in the blank questions, multiple choice questions, multiple select questions, and matching tables. The Tennessee TCAP assessment also uses a scaled score for student scoring.

## Population and Sample

The population included all third graders, all fourth graders, and all fifth graders within the state of Tennessee that utilized the Renaissance STAR and TCAP assessments. For this study the sample population included 484 third graders for Math and 604 third graders for Reading, 604 fourth graders for Math and 525 fourth graders for Reading, and 452 fifth graders for Math and 473 fifth graders for Reading, for a total of 3,142 subjects. The sample was selected based on the availability of schools using both testing platforms, a willingness to share students testing scores, and geographical location. All of the included testing scores are from students attending participating rural schools from two districts in Southeast Tennessee during the 2016-2017 school year. The researcher has not been granted access to personal identifying student data for sub-groups including gender and ethnicity.

## Data Collection

Existing test scores, ex post facto data, from two school administered assessments were collected by the researcher for the purpose of this study. Test scores from the school districts were collected by school administrators on behalf of the researcher. Security and privacy guidelines restrict direct access, by the researcher, to student test scores. In order to protect personal identifying student data, administrators at these participating schools compiled data based on research requests for test scores. Under instruction of the researcher, participating administrators matched student names and listed both test scores in an organized list. Individual student scores from the STAR and TCAP were paired. The prepared lists provided to the researcher were divided into three grade level sub-groups, third grade, fourth grade, and fifth grade. All identifying student information was excluded from the data the researcher received. Since it would be necessary to see student names in order to pair the two test scores for each participant, the researcher relied on school administrators to compile an accurate and reliable data set.

Due to the nature of the study, it was necessary to have paired data for Math and Reading respectively. A valid STAR and a TCAP score in each subject area was imperative to the research data collection process. Consequently, students with a missing score in either content area, Math or Reading were excluded from that portion of the research.

## Data Analysis

A series of Pearson correlations were computed to determine the relationship between the two variables. The researcher used the Statistical Program for the Social Sciences (IBM-SPSS) program to input p-values for each of the selected testing formats and ran a test to see whether the two variables have a statistically significant correlation and recorded the descriptive statics for the test score data sets. The variables were the scaled score from the STAR assessment and the scaled score from the TCAP assessment. The IBM-SPSS was used to analyze the data and all data was analyzed at the .05 level of significance.

A one-way analysis of variance was conducted to evaluate the relationship between the STAR test data and the TCAP test data. The factor variable, the STAR test data, included three levels: low which is known as urgent intervention, middle which is called intervention, and a high group which is labeled at/beyond benchmark. The dependent variable was the TCAP test scores. Since the overall *F* test was significant, post hoc multiple comparisons were used to evaluate pairwise differences among the means of the three groups.

## Chapter Summary

The researcher completed an ex post facto quantitative study on two different assessments used by school systems. The Renaissance STAR test is administered to students via computer three times per school year as a benchmark assessment, and TCAP is given once a year as it is the state of Tennessee's end of year assessment. The format of the state assessment changed for the 2016-2017 school year. No longer are students answering all multiple-choice questions on the TCAP assessment. The 2016-2017 TCAP assessment introduced third, fourth, and fifth graders to a combination of multiple-choice, multiple response, and short answer questions.

With the variations of the testing platforms, the Renaissance STAR test is administered on a computer and the TCAP is completed using paper and pencil, it was important to see if the results from the two tests correlate. In addition to the variances in the administration of the tests, the formatting of the items on the tests also varies. The Renaissance STAR tests are made up of 100% multiple-choice questions and the Tennessee end of year state assessment, TCAP, has a combination of multiple-choice, multiple response, and short response questions. It was important for me as a teacher to see if a correlation existed between the two tests for the purpose of planning and instruction for my students. If the two tests had not reflected a correlation, it might have been of interest to teachers and administrators to select and purchase a different testing platform more closely related to the state TCAP test.

# CHAPTER 4

# FINDINGS

This chapter contains the results of the data analysis as it relates to the twelve research questions proposed in Chapters 1 and 3. The purpose of this study was to correlate the student scores from the Renaissance Place STAR assessment to TCAP in Math and Reading and to evaluate the difference between TCAP scores when compared by STAR Reading or STAR Math level (Urgent Intervention, Intervention, At/Beyond Benchmark). The data were gathered from two standardized assessments in Math and Reading for third, fourth, and fifth grade students from two participating school districts in Southeastern Tennessee. The TCAP and the STAR scores were collected for the 2016-2017 school year. Chapter 4 was guided by twelve research questions and associated null hypotheses.

## Analysis of Research Questions

### Research Question 1

Is there a significant correlation between the Renaissance STAR Reading test scores and TCAP in Reading for students in 3rd grade?

$H_{01}$: There is not a significant correlation between the Renaissance STAR Reading test scores and TCAP in Reading for students in 3rd grade.

A Pearson correlation coefficient was computed to test the relationship between the STAR Reading scores and TCAP Reading scores for 3rd grade students. The results of the analysis revealed a significant positive relationship between STAR ($M = 483.41$, $SD = 170.285$) and TCAP ($M = 344.82$, $SD = 35.053$) [$r(603) = .669$, $p < .001$]. As a result of the analysis, the null hypothesis was rejected. In general, the results suggest that 3rd graders with high STAR scores in Reading also tended to have high TCAP scores in Reading. Figure 1 shows a scatterplot for the 3rd grade TCAP and STAR Reading scores.

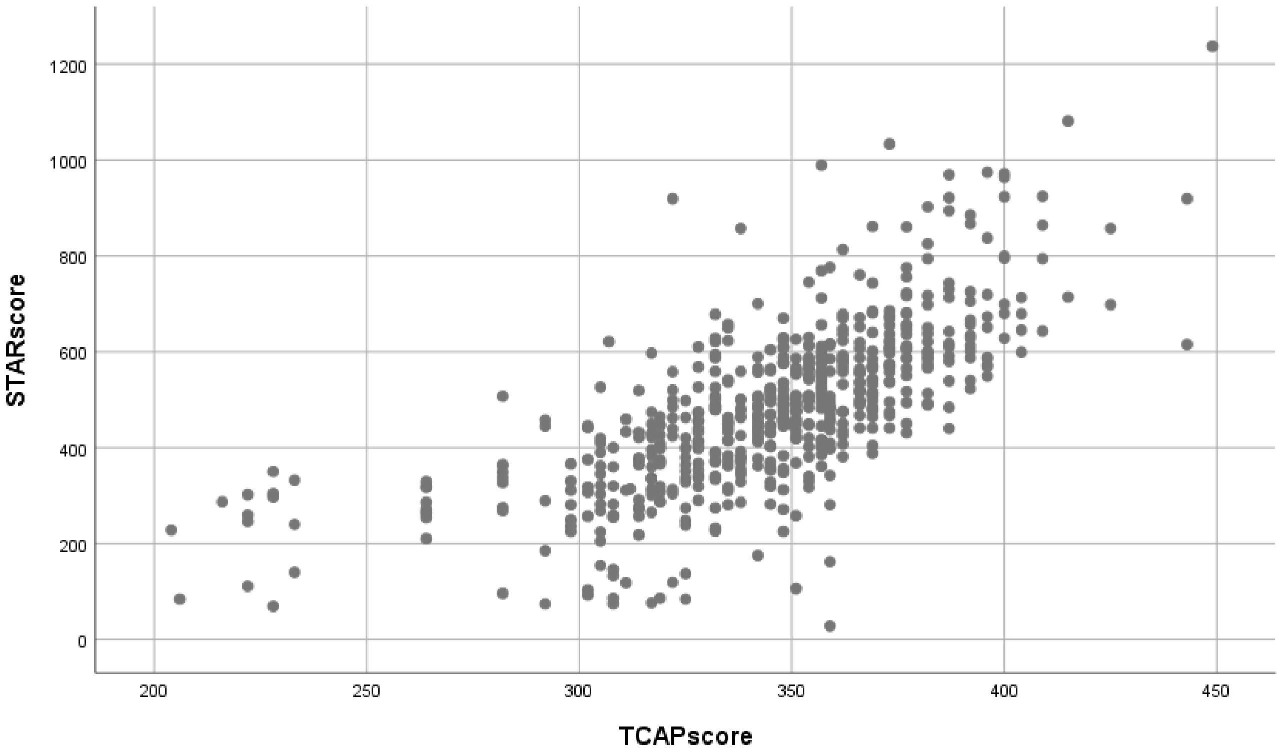



*Figure 1.* 3rd grade STAR and TCAP Reading scores

**Research Question 2**

Is there a significant correlation between the Renaissance STAR Reading test scores and TCAP in Reading for students in 4th grade?

$H_{02}$: There is not a significant correlation between the Renaissance STAR Reading test scores and TCAP in Reading for students in 4th grade.

A Pearson correlation coefficient was computed to test the relationship between the STAR Reading scores and TCAP Reading scores for 4th grade students. The results of the analysis revealed a significant positive relationship between STAR (M = 599.38, SD = 189.957) and TCAP (M = 333.27, SD = 30.349) [r(524) = .688, p < .001]. As a result of the analysis, the null hypothesis was rejected. In general, the results suggest that 4th graders with high STAR scores in Reading also tended to have high TCAP scores in Reading. Figure 2 shows a scatterplot for the 4th grade TCAP and STAR Reading scores.

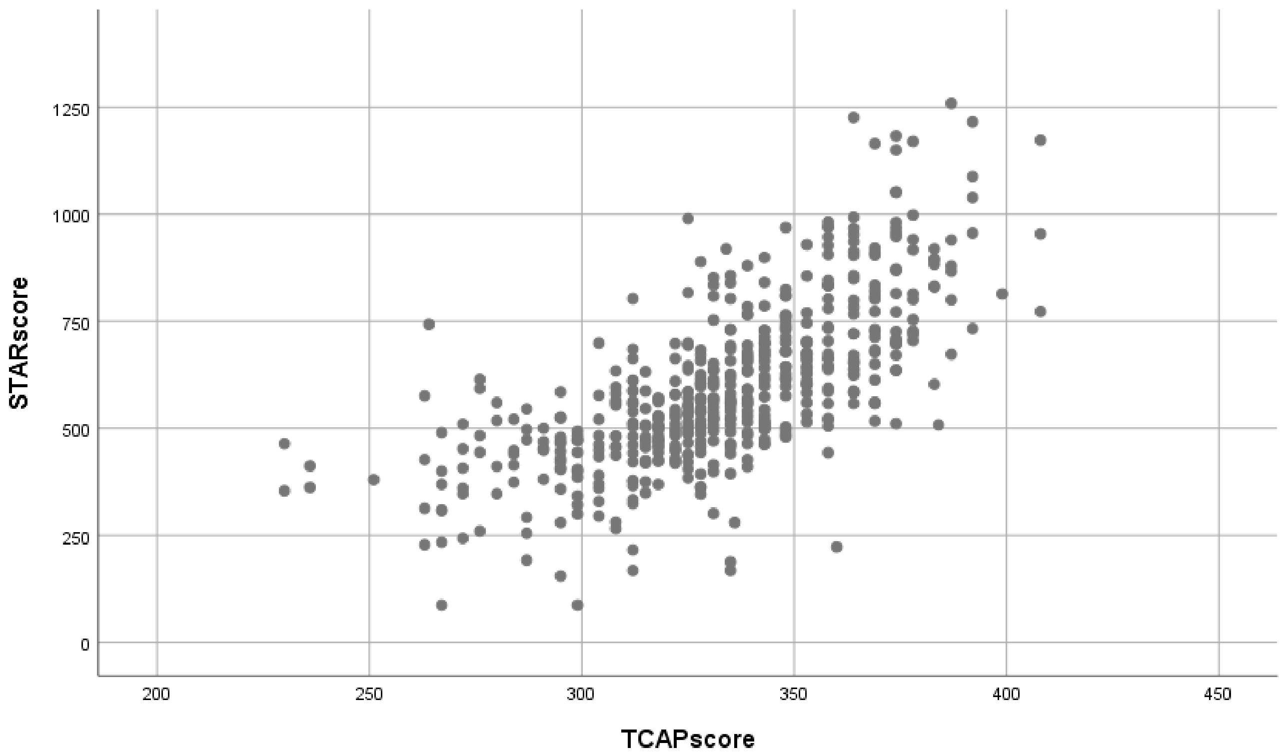



*Figure 2.* 4th grade STAR and TCAP Reading scores

**Research Question 3**

Is there a significant correlation between the Renaissance STAR Reading test scores and TCAP in Reading for students in 5th grade?

$H_{03}$: There is not a significant correlation between the Renaissance STAR Reading test scores and TCAP in Reading for students in 5th grade.

A Pearson correlation coefficient was computed to test the relationship between the STAR Reading scores and TCAP Reading scores for 5th grade students. The results of the analysis revealed a significant positive relationship between STAR (M = 672.27, SD = 213.056) and TCAP (M = 317.21, SD = 29.927) [r(472) = .678, p < .001]. As a result of the analysis, the null hypothesis was rejected. In general, the results suggest that 5th graders with high STAR scores in Reading also tended to have high TCAP scores in Reading. Figure 3 shows a scatterplot for the 5th grade TCAP and STAR Reading scores.

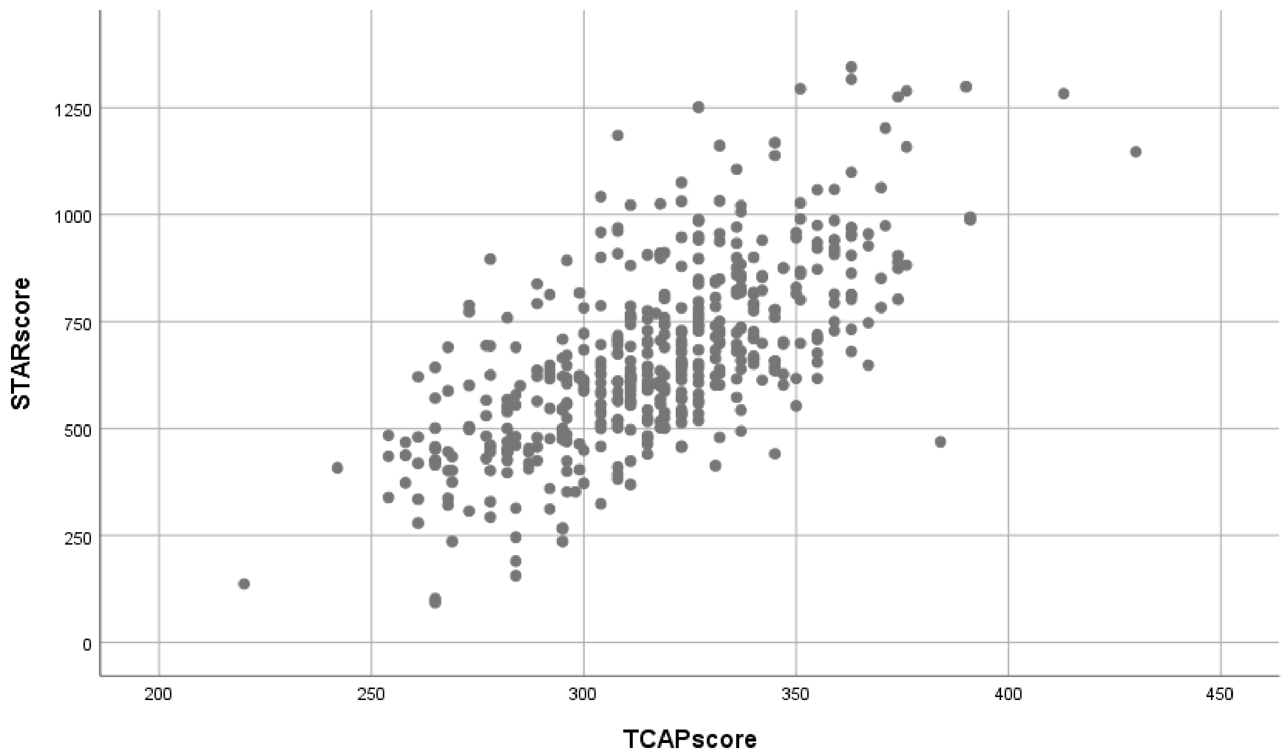



*Figure 3.* 5th grade STAR and TCAP Reading scores

**Research Question 4**

Is there a significant correlation between the Renaissance STAR Math test scores and TCAP in Math for students in 3rd grade?

$H_{04}$: There is not a significant correlation between the Renaissance STAR Math test scores and TCAP in Math for students in 3rd grade.

A Pearson correlation coefficient was computed to test the relationship between the STAR Math scores and TCAP Math scores for 3rd grade students. The results of the analysis revealed a significant positive relationship between STAR (M = 633.16, SD = 75.854) and TCAP (M = 333.70, SD = 37.241) [r(483) = .780, p < .001]. As a result of the analysis, the null hypothesis was rejected. In general, the results suggest that 3rd graders with high STAR scores in Math also tended to have high TCAP scores in Math. Figure 4 shows a scatterplot for the 3rd grade TCAP and STAR Math scores.

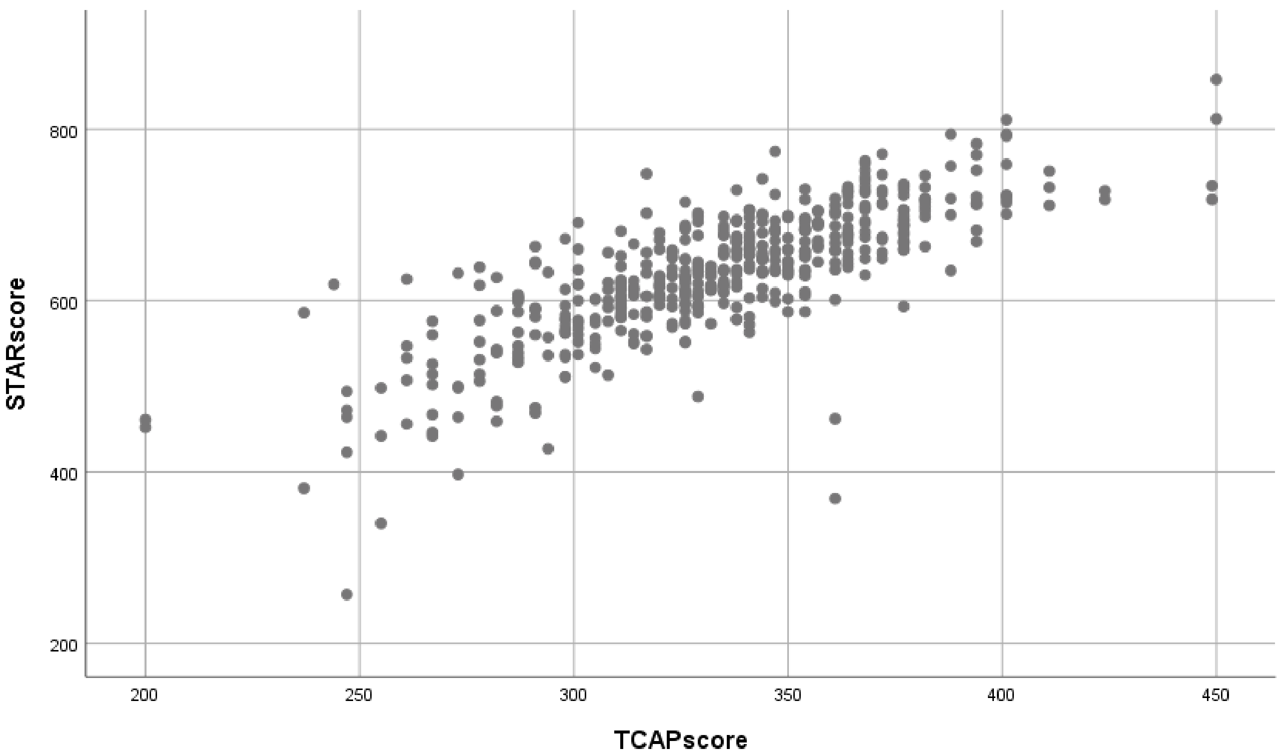



*Figure 4.* 3rd grade STAR and TCAP Math scores

**Research Question 5**

Is there a significant correlation between the Renaissance STAR Math test scores and TCAP in Math for students in 4th grade?

$H_{05}$: There is not a significant correlation between the Renaissance STAR Math test scores and TCAP in Math for students in 4th grade.

A Pearson correlation coefficient was computed to test the relationship between the STAR Math scores and TCAP Math scores for 4th grade students. The results of the analysis revealed a significant positive relationship between STAR (M = 691.07, SD = 110.633) and TCAP (M = 319.23, SD = 37.099) [r(603) = .635, p < .001]. As a result of the analysis, the null hypothesis was rejected. In general, the results suggest that 4th graders with high STAR scores in Math also tended to have high TCAP scores in Math. Figure 5 shows a scatterplot for the 4th grade TCAP and STAR Math scores.

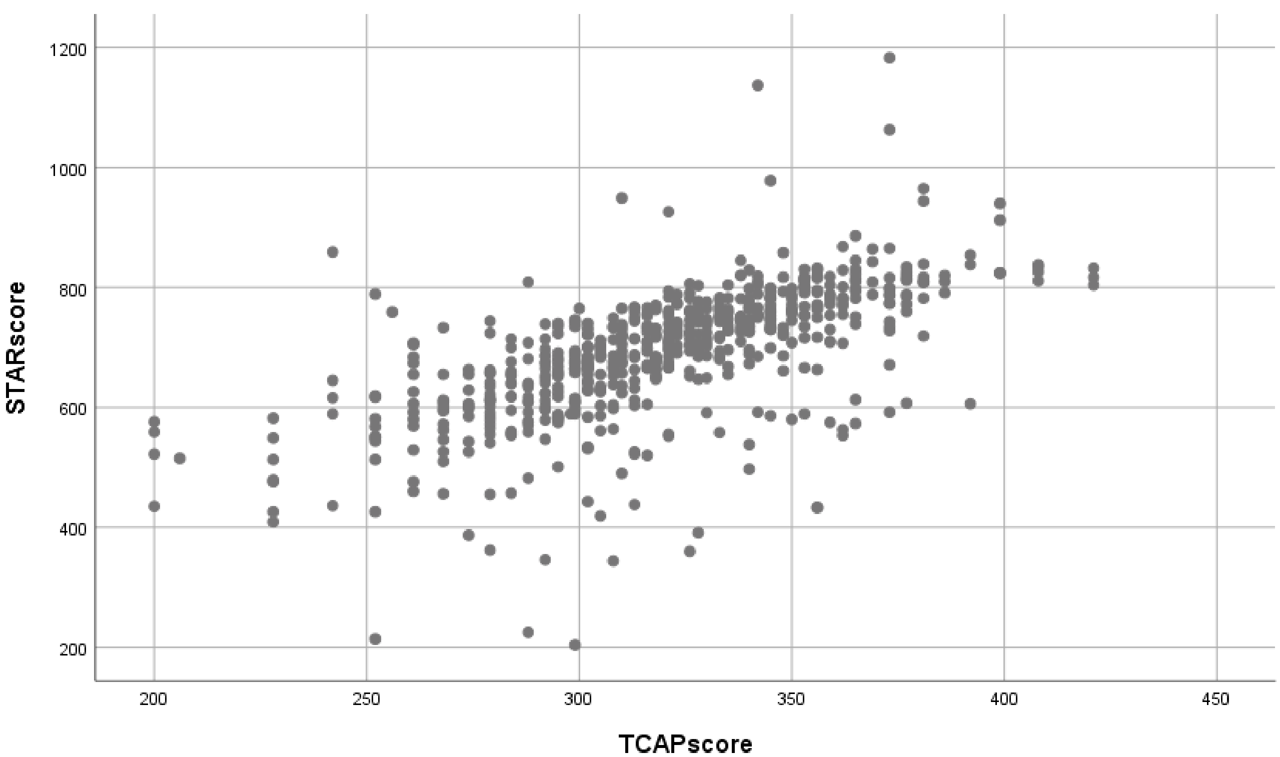



*Figure 5.* 4th grade STAR and TCAP Math scores

**Research Question 6**

Is there a significant correlation between the Renaissance STAR Math test scores and TCAP in Math for students in 5th grade?

$H_{06}$: There is not a significant correlation between the Renaissance STAR Math test scores and TCAP in Math for students in 5th grade.

A Pearson correlation coefficient was computed to test the relationship between the STAR Math scores and TCAP Math scores for 5th grade students. The results of the analysis revealed a significant positive relationship between STAR ($M = 728.42$, $SD = 89.593$) and TCAP ($M = 307.53$, $SD = 38.218$) [$r(451) = .803$, $p < .001$]. As a result of the analysis, the null hypothesis was rejected. In general, the results suggest that 5th graders with high STAR scores in Math also tended to have high TCAP scores in Math. Figure 6 shows a scatterplot for the 4th grade TCAP and STAR Math scores.

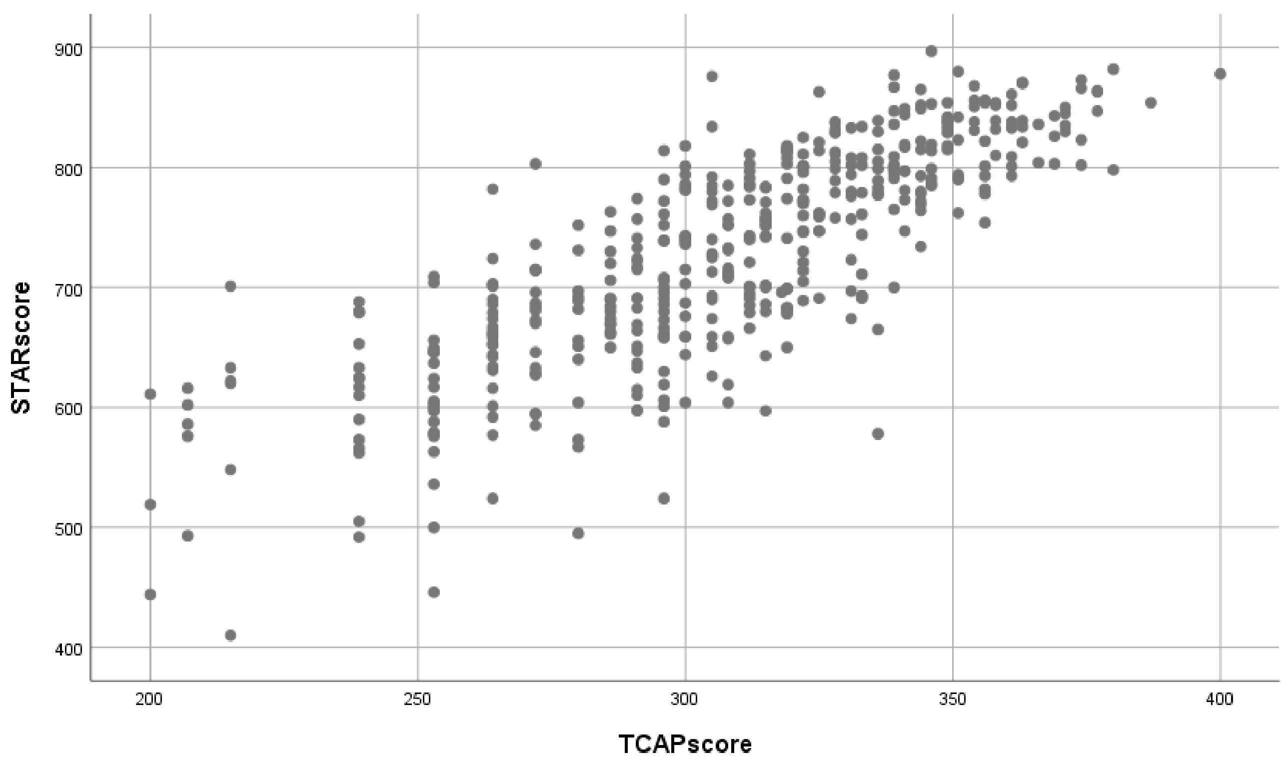



*Figure 6*. 5th grade STAR and TCAP Math scores

**Research Question 7**

Is there a statistically significant difference in 3rd grade TCAP Reading scores as compared by STAR Reading level (Urgent Intervention, Intervention, At/Beyond Benchmark)?

$H_{07}$: There is not a statistically significant difference in 3rd grade TCAP Reading scores as compared by STAR Reading level (Urgent Intervention, Intervention, At/Beyond Benchmark).

A one-way analysis of variance was conducted to evaluate the relationship between the percentile category of the STAR test and the TCAP test. The factor variable, the percentile category, included three levels: Urgent Intervention, Intervention, and At/Beyond Benchmark. The dependent variable was the TCAP reading score. The ANOVA was significant, $F(2, 601) = 150.140$, $p < .001$. Therefore, the null hypothesis was rejected. The strength of the relationship between the STAR percentile group and the TCAP score, as assessed by $\eta^2$, was large (.333).

Because the overall *F* test was significant, post hoc multiple comparisons were conducted to evaluate pairwise difference among the means of the three groups. A Tukey procedure was selected for the multiple comparisons because equal variances were assumed. Third grade TCAP Reading scores from the At/Beyond Benchmark group were significantly higher than both the Urgent Intervention group ($p < .001$) and the Intervention group ($p < .001$). However, there was not a significant difference between the Urgent Intervention group and the Intervention group ($p = .236$). The 95% confidence intervals for the pairwise differences, as well as, the means and standard deviations for the three STAR percentile groups, are reported in Table 1. Figure 7 shows the pairwise differences for 3rd grade Reading.

Table 1
*Means, Standard Deviations, and Confidence Intervals of Pairwise Differences 3rd grade Reading TCAP Scores*

| Percentile Category | N | M | SD | Urgent Intervention | Intervention |
|---|---|---|---|---|---|
| Urgent Intervention | 46 | 297.78 | 41.148 | | |
| Intervention | 68 | 306.68 | 33.263 | -3.97-21.75 | |
| At/Beyond Benchmark | 490 | 354.52 | 26.514 | 46.35-67.13 | 39.13-56.57 |

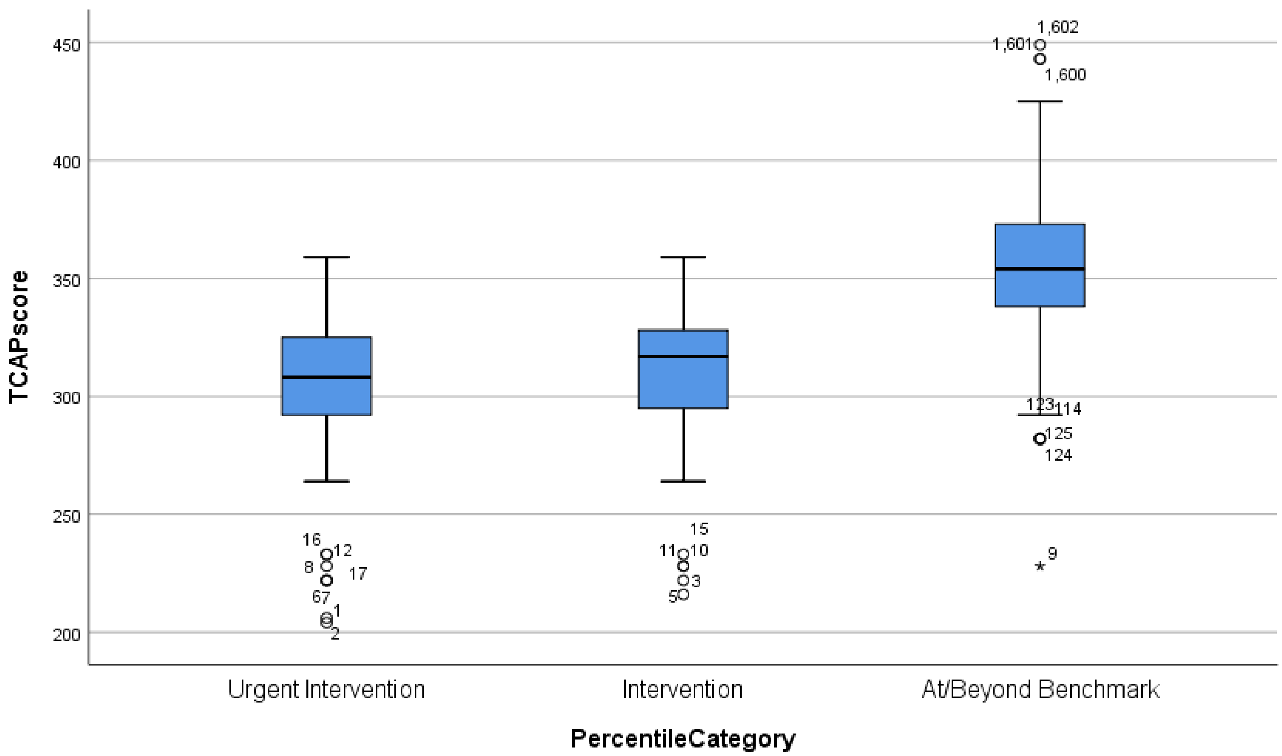



*Figure 7.* Pairwise Differences 3rd grade Reading

**Research Question 8**

Is there a statistically significant difference in 4th grade TCAP Reading scores as compared by STAR Reading level (Urgent Intervention, Intervention, At/Beyond Benchmark)?

$H_{08}$: There is not a statistically significant difference in 4th grade TCAP Reading scores as compared by STAR Reading level (Urgent Intervention, Intervention, At/Beyond Benchmark).

A one-way analysis of variance was conducted to evaluate the relationship between the percentile category of the STAR test and the TCAP test. The factor variable, the percentile category, included three levels: Urgent Intervention, Intervention, and At/Beyond Benchmark. The dependent variable was the TCAP reading score. The ANOVA was significant, $F(2, 522) = 66.149$, $p < .001$. Therefore, the null hypothesis was rejected. The strength of the relationship between the STAR percentile group and the TCAP score, as assessed by $\eta^2$, was large (.202).

Because the overall *F* test was significant, post hoc multiple comparisons were conducted to evaluate pairwise difference among the means of the three groups. A Tukey procedure was selected for the multiple comparisons because equal variances were assumed. Fourth grade TCAP Reading scores from the At/Beyond Benchmark group were significantly higher than both the Urgent Intervention group ($p < .001$) and the Intervention group ($p < .001$). However, there was not a significant difference between the Urgent Intervention group and the Intervention group ($p = .998$). The 95% confidence intervals for the pairwise differences, as well as, the means and standard deviations for the three STAR percentile groups, are reported in Table 2. Figure 8 shows the pairwise differences for 4th grade Reading.

Table 2
*Means, Standard Deviations, and Confidence Intervals of Pairwise Differences 4th grade Reading TCAP Scores*

| Percentile Category | N | M | SD | Urgent Intervention | Intervention |
|---|---|---|---|---|---|
| Urgent Intervention | 27 | 297.89 | 26.148 | | |
| Intervention | 40 | 297.45 | 27.959 | -16.34 – 15.46 | |

| | | | | | |
|---|---|---|---|---|---|
| At/Beyond Benchmark | 458 | 338.49 | 27.147 | 27.96 – 53.24 | 30.51 – 51.56 |

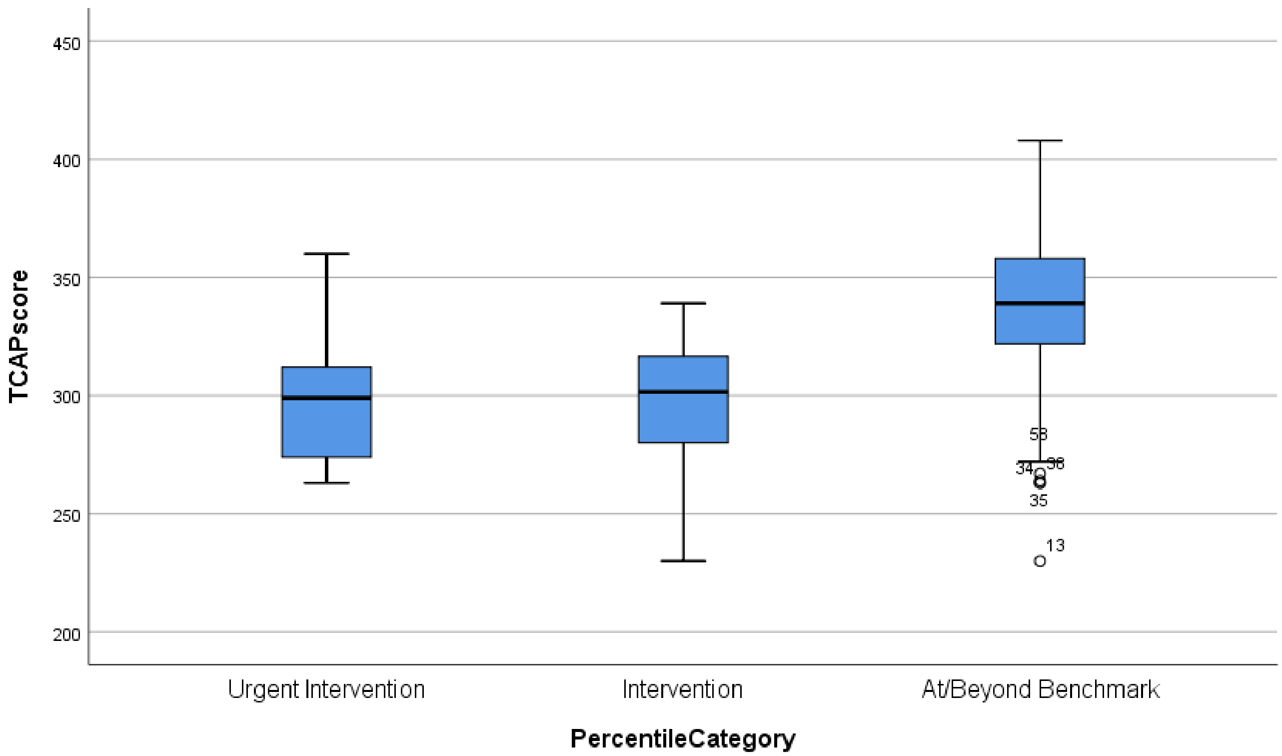



*Figure 8.* Pairwise Differences 4th grade Reading

**Research Question 9**

Is there a statistically significant difference in 5th grade TCAP Reading scores as compared by STAR Reading level (Urgent Intervention, Intervention, At/Beyond Benchmark)?

$H_{09}$: There is not a statistically significant difference in 5th grade TCAP Reading scores as compared by STAR Reading level (Urgent Intervention, Intervention, At/Beyond Benchmark).

A one-way analysis of variance was conducted to evaluate the relationship between the percentile category of the STAR test and the TCAP test. The factor variable, the percentile category, included three levels: Urgent Intervention, Intervention, and At/Beyond Benchmark.

The dependent variable was the TCAP reading score. The ANOVA was significant, F(2, 470) = 88.829, p < .001. Therefore, the null hypothesis was rejected. The strength of the relationship between the STAR percentile group and the TCAP score, as assessed by $\eta^2$, was large (.274).

Because the overall *F* test was significant, post hoc multiple comparisons were conducted to evaluate pairwise difference among the means of the three groups. A Tukey procedure was selected for the multiple comparisons because equal variances were assumed. Fifth grade TCAP Reading scores from the At/Beyond Benchmark group were significantly higher than both the Urgent Intervention group (p < .001) and the Intervention group (p < .001). However, there was not a significant difference between the Urgent Intervention group and the Intervention group (*p* = .317). The 95% confidence intervals for the pairwise differences, as well as, the means and standard deviations for the three STAR percentile groups, are reported in Table 3. Figure 9 shows the pairwise differences for 5th grade Reading.

Table 3
*Means, Standard Deviations, and Confidence Intervals of Pairwise Differences 5th grade Reading TCAP Scores*

| Percentile Category | N | M | SD | Urgent Intervention | Intervention |
|---|---|---|---|---|---|
| Urgent Intervention | 30 | 280.57 | 20.018 | | |
| Intervention | 65 | 288.74 | 24.901 | -5.09 - 21.43 | |
| At/beyond benchmark | 378 | 325.01 | 26.031 | 33.05 – 55.83 | 28.20 – 44.34 |

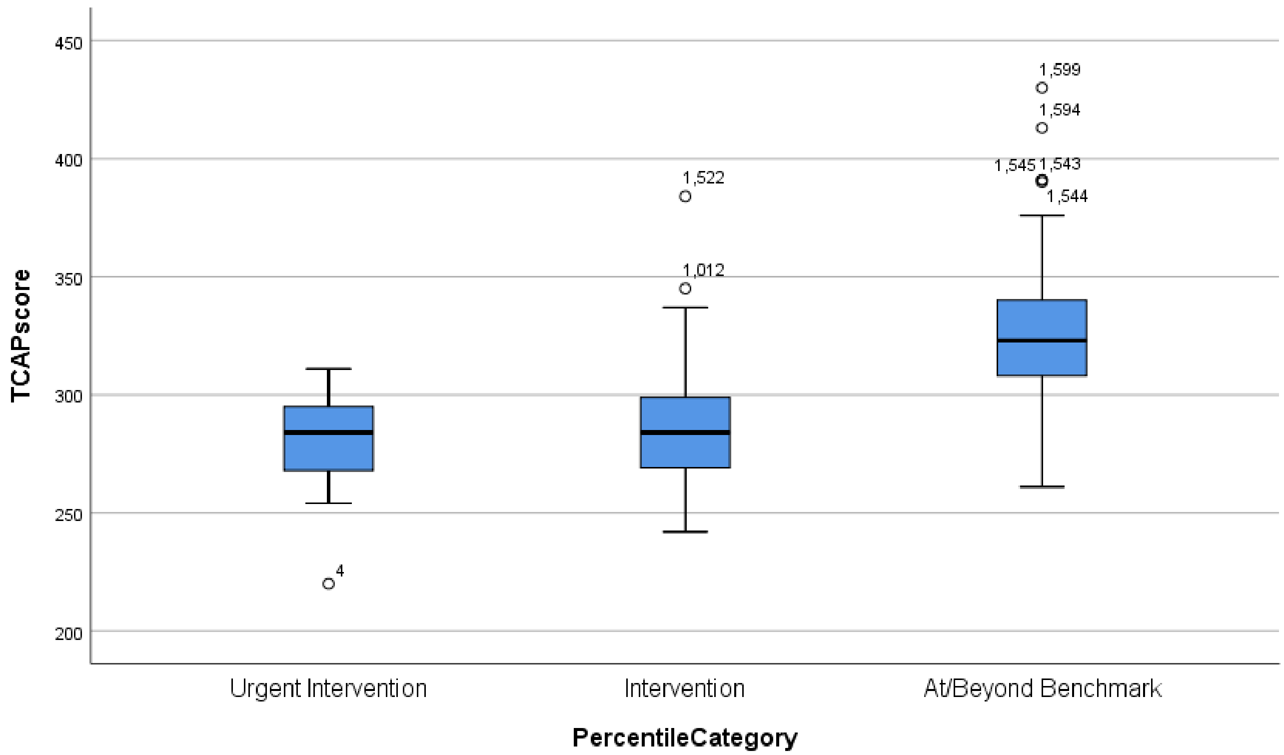



*Figure 9*. Pairwise Differences 5th grade Reading

**Research Question 10**

Is there a statistically significant difference in 3rd grade TCAP Math scores as compared by STAR Math level (Urgent Intervention, Intervention, At/Beyond Benchmark)?

$H_{010}$: There is not a statistically significant difference in 3rd grade TCAP Math scores as compared by STAR Math level (Urgent Intervention, Intervention, At/Beyond Benchmark).

A one-way analysis of variance was conducted to evaluate the relationship between the percentile category of the STAR test and the TCAP test. The factor variable, the percentile category, included three levels: Urgent Intervention, Intervention, and At/Beyond Benchmark. The dependent variable was the TCAP reading score. The ANOVA was significant, $F(2, 481) = 83.392$, $p < .001$. Therefore, the null hypothesis was rejected. The strength of the relationship between the STAR percentile group and the TCAP score, as assessed by $\eta^2$, was large (.257).

Because the overall *F* test was significant, post hoc multiple comparisons were conducted to evaluate pairwise difference among the means of the three groups. A Tukey procedure was selected for the multiple comparisons because equal variances were assumed. Third grade TCAP Math scores from the At/Beyond Benchmark group were significantly higher than both the Urgent Intervention group ($p < .001$) and the Intervention group ($p < .001$). However, there was not a significant difference between the Urgent Intervention group and the Intervention group ($p = .193$). The 95% confidence intervals for the pairwise differences, as well as, the means and standard deviations for the three STAR percentile groups, are reported in Table 4. Figure 10 shows the pairwise differences for 3rd grade Math.

Table 4
*Means, Standard Deviations, and Confidence Intervals of Pairwise Differences 3rd grade Math TCAP Scores*

| Percentile Category | N | M | SD | Urgent Intervention | Intervention |
|---|---|---|---|---|---|
| Urgent Intervention | 20 | 266.60 | 40.440 | | |
| Intervention | 26 | 283.19 | 18.717 | -5.89 – 39.08 | |
| At/beyond benchmark | 438 | 339.76 | 32.359 | 55.88 – 90.45 | 41.31 – 71.83 |

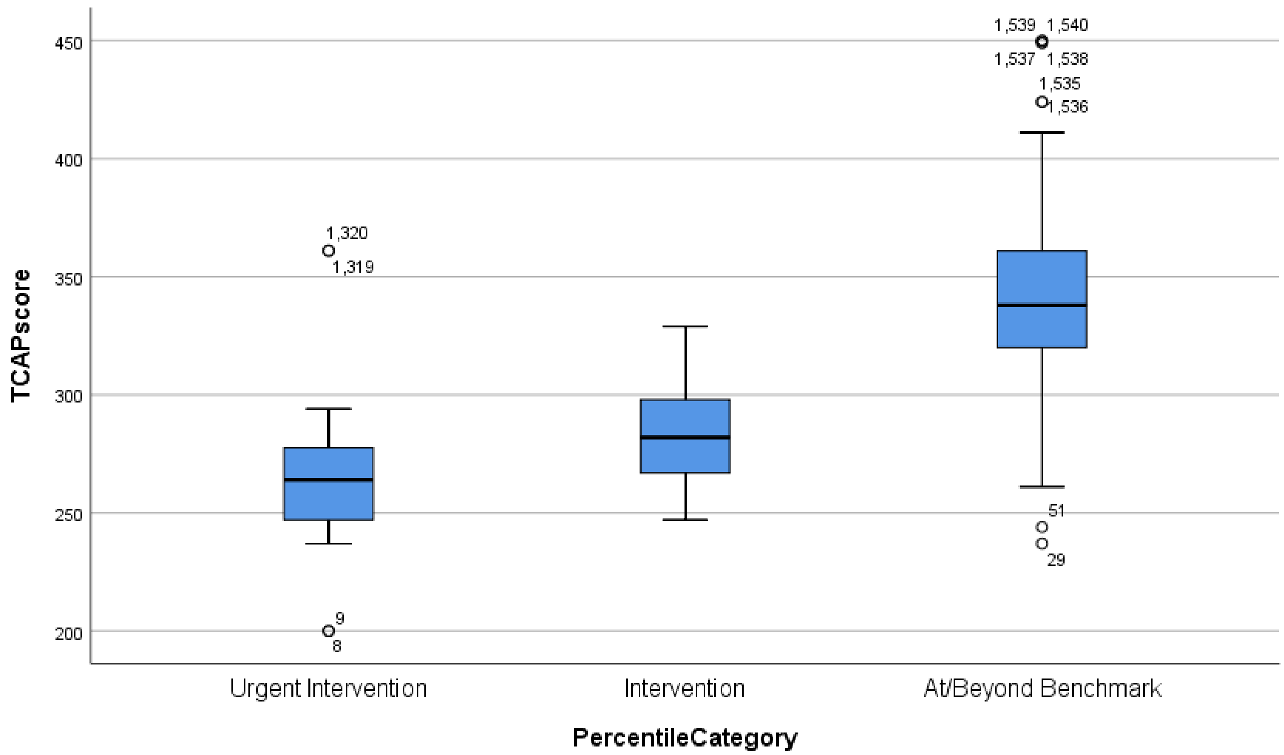



*Figure 10.* Pairwise Differences 3rd grade Math

**Research Question 11**

Is there a statistically significant difference in 4th grade TCAP Math scores as compared by STAR Math level (Urgent Intervention, Intervention, At/Beyond Benchmark)?

$H_{011}$: There is not a statistically significant difference in 4th grade TCAP Math scores as compared by STAR Math level (Urgent Intervention, Intervention, At/Beyond Benchmark).

A one-way analysis of variance was conducted to evaluate the relationship between the percentile category of the STAR test and the TCAP test. The factor variable, the percentile category, included three levels: Urgent Intervention, Intervention, and At/Beyond Benchmark. The dependent variable was the TCAP reading score. The ANOVA was significant, $F(2, 601) = 65.598$, $p < .001$. Therefore, the null hypothesis was rejected. The strength of the relationship between the STAR percentile group and the TCAP score, as assessed by $\eta^2$, was medium (.179).

Because the overall *F* test was significant, post hoc multiple comparisons were conducted to evaluate pairwise difference among the means of the three groups. A Tukey procedure was selected for the multiple comparisons because equal variances were assumed. Fourth grade TCAP Math scores from the At/Beyond Benchmark group were significantly higher than both the Urgent Intervention group ($p < .001$) and the Intervention group ($p < .001$). However, there was not a significant difference between the Urgent Intervention group and the Intervention group ($p = .288$). The 95% confidence intervals for the pairwise differences, as well as, the means and standard deviations for the three STAR percentile groups, are reported in Table 5. Figure 11 shows the pairwise differences for 4th grade Math.

Table 5
*Means, Standard Deviations, and Confidence Intervals of Pairwise Differences 4th grade Math TCAP Scores*

| Percentile Category | N | M | SD | Urgent Intervention | Intervention |
|---|---|---|---|---|---|
| Urgent Intervention | 44 | 279.84 | 40.217 | | |
| Intervention | 63 | 289.81 | 37.447 | -5.57 – 25.51 | |
| At/beyond benchmark | 497 | 326.45 | 32.526 | 34.17 – 59.05 | 26.06 – 47.22 |

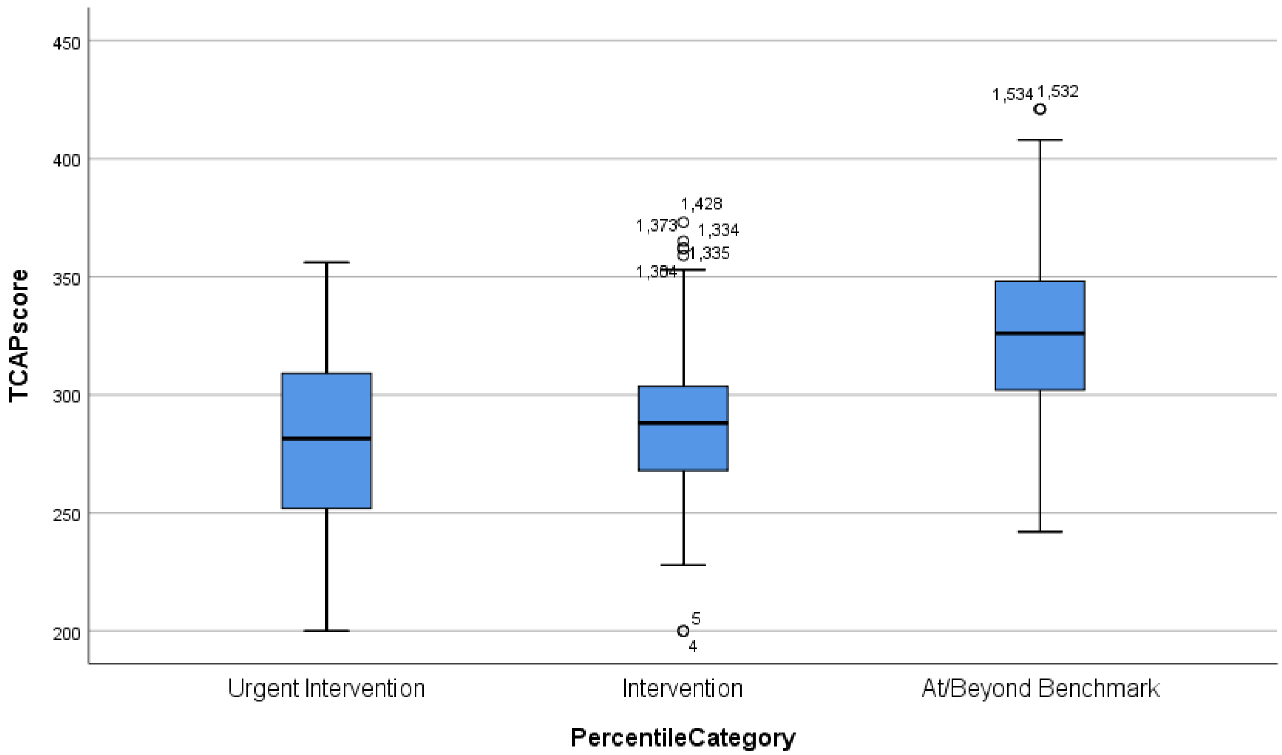



*Figure 11.* Pairwise Differences 4th grade Math

**Research Question 12**

Is there a statistically significant difference in 5th grade TCAP Math scores as compared by STAR Math level (Urgent Intervention, Intervention, At/Beyond Benchmark)?

$H_{012}$: There is not a statistically significant difference in 5th grade TCAP Math scores as compared by STAR Math level (Urgent Intervention, Intervention, At/Beyond Benchmark).

A one-way analysis of variance was conducted to evaluate the relationship between the percentile category of the STAR test and the TCAP test. The factor variable, the percentile category, included three levels: Urgent Intervention, Intervention, and At/Beyond Benchmark. The dependent variable was the TCAP reading score. The ANOVA was significant, $F(2, 449) = 131.424$, $p < .001$. Therefore, the null hypothesis was rejected. The strength of the relationship between the STAR percentile group and the TCAP score, as assessed by $\eta^2$, was large (.369).

Because the overall *F* test was significant, post hoc multiple comparisons were conducted to evaluate pairwise difference among the means of the three groups. A Tukey procedure was selected for the multiple comparisons because equal variances were assumed. Fifth grade TCAP Math scores from the At/Beyond Benchmark group were significantly higher than both the Urgent Intervention group ($p < .001$) and the Intervention group ($p < .001$). In addition, the 5th grade TCAP Math scores from the Intervention group were significantly higher than the Urgent Intervention group ($p = .039$). The 95% confidence intervals for the pairwise differences, as well as, the means and standard deviations for the three STAR percentile groups, are reported in Table 6. Figure 12 shows the pairwise differences for 5th grade Math.

Table 6
*Means, Standard Deviations, and Confidence Intervals of Pairwise Differences 5th grade Math TCAP Scores*

| Percentile Category | N | M | SD | Urgent Intervention | Intervention |
|---|---|---|---|---|---|
| Urgent Intervention | 30 | 249.23 | 31.613 | | |
| Intervention | 56 | 266.11 | 28.784 | 0.69– 33.06 | |
| At/beyond benchmark | 366 | 318.53 | 30.563 | 55.82 – 82.99 | 42.27 – 62.80 |

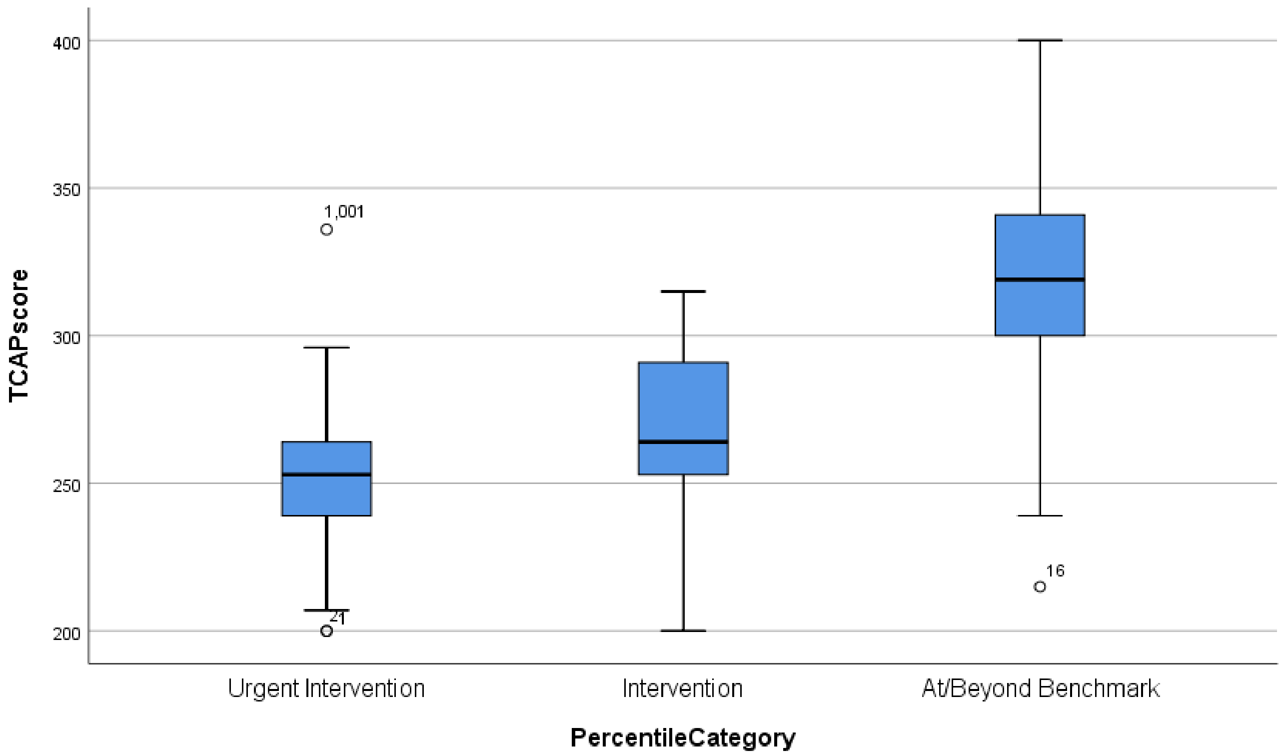



*Figure 12.* Pairwise Differences 5th grade Math

## Chapter Summary

In this chapter STAR and TCAP assessments scores used for analysis were collected from the 2016-2017 school year for grades 3, 4, and 5. School administrators paired individual student STAR scores with individual student TCAP scores. The complied list of scores was provided to the researcher for data analysis.

There were six research questions and six null hypotheses related to correlation. A series of Pearson correlation coefficient were computed to test the relationship between the STAR Reading scores and TCAP Reading scores for 3rd, 4th, and 5th grade students in Math and Reading. In general, the results suggest that students with high STAR scores in Reading also tended to have high TCAP scores in Reading and students with high STAR scores in Math also tended to have high TCAP scores in Math.

There were six research questions and six null hypotheses related to the evaluation of the relationship between the percentile category of the STAR test and the TCAP test. A series of one-way analysis of variance were conducted to evaluate the relationship between the percentile category of the STAR test and the TCAP test. The factor variable, the percentile category, included three levels: Urgent Intervention, Intervention, and At/Beyond Benchmark. The dependent variable was the TCAP reading score. In general, the results suggested TCAP scores in Math and Reading from the At/Beyond Benchmark group were significantly higher than both the Urgent Intervention group and the Intervention group.

# CHAPTER 5

# SUMMARY, CONCLUSION, AND RECOMMENDATIONS

The purpose of this study was twofold. One objective was to determine if there was a significant correlation between the Renaissance STAR test scores and TCAP in Math and Reading for students in 3rd, 4th, and 5th grade. The second purpose of the study was to test for a statistically significant difference in TCAP scores when compared by STAR level (Urgent Intervention, Intervention, At/Beyond Benchmark) in Math and Reading for 3rd, 4th, and 5th grade students.

## Summary of Results

A series of Pearson correlations were computed to test the relationship between the STAR scores and TCAP scores in Math and Reading for students in 3rd, 4th, and 5th grade. The results of the analyses revealed a significant positive relationship between STAR assessment and the TCAP assessment in Math and Reading for grades 3, 4, and 5. In general, the results suggest that 3rd, 4th, and 5th graders with high STAR scores in Math also tended to have high TCAP scores in Math, and 3rd, 4th, and 5th graders with high STAR scores in Reading also tended to have high TCAP scores in Reading.

The results from this study contradict Risko and Walker-Dalhouse (2010), when they asserted most benchmark tests are not useful in relation to instructional decision making because these tests are not directly related to educational benefits for students. Wiliam (2017) found the use of assessment during the learning process as opposed to an assessment at the completion of teaching, will have a greater impact on how quickly students learn than almost any other factor.

Based on the results of this study, teachers can use the STAR test as formative assessment to guide instruction to maximize students learning in Math and Reading for student in grade 3, 4, and 5.

Within the second portion of this study the researcher tested for statistically significant differences in TCAP scores when compared by STAR level (Urgent Intervention, Intervention, At/Beyond Benchmark) in Math and Reading for 3rd, 4th, and 5th grade students. The factor variable, the percentile category from the STAR assessment, included three levels: Urgent Intervention, Intervention, and At/Beyond Benchmark. The dependent variable was the TCAP Math or Reading score. Results indicated a strong relationship between the STAR percentile group and the TCAP score.

When analyzed, the TCAP scores from the At/Beyond Benchmark group were significantly higher than both the Urgent Intervention group and the Intervention group in Math and Reading for students in 3rd, 4th and 5th grade. The students scoring in the At/Beyond Benchmark category on the STAR assessment were performing significantly higher on the TCAP assessment than their peers who scored in the Urgent Intervention or Intervention category on the STAR assessment. A statistically significant difference was unfounded on the TCAP assessment when comparing the Urgent Intervention and Intervention groups from the STAR assessment, with the sole exception being 5$^{th}$ grade Math. The 5th grade TCAP Math scores from the Intervention group were significantly higher than the Urgent Intervention group.

Marzano (2010) suggested formative assessments be used to evaluate student learning during the lesson to allows teachers the ability to make instructional adjustments to enhance student learning. Based on the information founded in this study, the STAR assessment can be used by teachers to drive instruction in the classroom in preparation for the TCAP assessment.

Consistent with the RTI Action Network (2017) in suggesting at-risk should receive additional support and supplemental materials in the regular classroom during the school day, this study revealed students scoring in the Urgent Intervention and Intervention group on the STAR are performing at a significantly lower rate on the TCAP than their peers

## Recommendations for Practice

Based on the findings of this study, the author provides the following recommendations for practice:

1. Educational assessment should be used in a manner that will positively impact the learning opportunities for students. A formalized assessment such as the STAR test should be used as a diagnostic tool for specific intervention to enhance student achievement.
2. Teachers and administrators should evaluate the implications of excessive testing. If teachers are not testing in a manner that will impact instruction within the classroom, it is the responsibility of the teacher to determine of the assessment is necessary.
3. Teachers and administrators should continue the use of the STAR Assessment within the classroom as a form of progress monitoring, a screener for RTI program, and too to assist in intervention instruction.
4. Due to the strong correlation with TCAP assessment, administrators and teachers in districts not using STAR may want to review the findings of this study and consider the potential benefits of this assessment. STAR does provide teachers with individualized student feedback in a timely manner which allows adjustments to classroom instruction.

5. Teachers and administrators from all school systems in Tennessee should consider the use of the STAR assessment as an instructional tool. The STAR assessment can accurately provide teachers and administrators with information related to individual student progress and individual educational advancement toward state standards as defined by the TCAP assessment.
6. Students scoring in the Urgent Intervention and Intervention groups on the STAR assessment should receive rigorous personalized support in order to better serve individual needs in preparation of the TCAP assessment.

## Recommendations for Further Research

The results of this study have prompted the researcher to make the following recommendations for further research regarding student achievement on the STAR assessment and the TCAP assessment:

1. Similar studies should be conducted in different grade levels, regions, or states with a larger sample size to determine if the results are similar to the results of this study.
2. Conduct research tracking individual student achievement on the STAR and TCAP over several years to determine if a pattern exists.
3. Conduct the same study over several years correlating and comparing STAR and TCAP scores for students in grades three, four, and five to determine if the STAR assessment is a predictor of student achievement on the TCAP.
4. Investigate the difference between student growth scores on the STAR assessment and student growth scores on the TCAP assessment over a number of years to determine is a relationship exists.

5. Conduct research to determine if student gender impacts assessment results on STAR when correlated and compared to TCAP.
6. Investigate the disproportion academic performance of sub-groups, identified as: racial/ethnic, socioeconomic, home-language, and special education groups on the STAR and TCAP assessments.
7. Use a qualitative or mixed methods approach to evaluate teacher, student, parent, administrator, and district leader perceptions of relationships between the STAR and the TCAP assessment.
8. Conduct a study that evaluates variations in prescribed RTI interventions by teacher and by school. By comparing differences in intervention practices used for RTI, the programs or methods that produce better results can be identified and utilized by teachers and administrators.

## Conclusion

An important aspect of testing as indicated by Haladyna et al. (2008), standardized testing should be used to identifying potential opportunities where intervention might benefit the individual. Due to the strong correlation between the STAR assessment and the TCAP assessment, classroom teachers should utilize the STAR assessment as a formative assessment tool. Individual student results from the STAR assessment should be analyzed by teachers for the purpose of providing instruction. The findings from this study are consistent with Vanderheyden (2013) describing a universal screener as a forecast of a child's performance on the end of year state assessment.

Students categorized by the STAR assessment as At/Beyond Benchmark are scoring significantly higher than their peers that score Urgent Intervention or Intervention on the STAR assessment. Scoring within the Urgent Intervention or Intervention on the STAR assessment should alert teachers to the critical instructional needs of a student. Students scoring Urgent Intervention or Intervention on the STAR are at a higher risk of low performance on the TCAP than their peers. The RTI Action Network (2017) suggested students identified as at-risk should receive additional support and supplemental materials in the regular classroom during the school day. Teachers should utilize all available resources to meet the urgent needs of students performing within the Urgent Intervention or Intervention on the STAR.

REFERENCES

Bennett, R. E. (2011). Formative assessment: A critical review. *Assessment in Education: Principles, Policy & Practice*, 18(1), 5-25. DOI: 10.1080/0969594X.2010.513678

Bennett, R. E., Braswell, J., Oranje, A., Sandene, B., Kaplan, B., & Yan, F. (2008). Does it matter if I take my mathematics test on computer? A second empirical study of mode effects in NAEP. *Journal of Technology, Learning, and Assessment*, 6(9), 1-39. Retrieved on November 4, 2017 from: https://ejournals.bc.edu/ojs/index.php/jtla/article/view/1639/1472

Birenbaum, M. & Feldman, R. A. (1998). Relationships between learning patterns and attitudes towards two assessment formats. *Educational Research,40*(1), 90-98. DOI: 10.1080/0013188980400109

Black, P., & Wiliam, D. (1998). Assessment and Classroom Learning. *Assessment in Education: Principles, Policy & Practice, 5*(1), 7-74. DOI: 10.1080/0969595980050102

Black, P. & Wiliam, D. (2003). 'In praise of educational research': Formative assessment. *British Educational Research Journal, 29*(5), 623-637. DOI: 10.1080/0141192032000133721

Black, P. & Wiliam, D. (2010). Inside the black box: Raising standards through classroom assessment. Phi Delta Kappan, 92(1), 81 – 90. DOI: 10.1177/ 003172171009200119

Bodmann, S. & Robinson, D. (2004). Speed and performance differences among computer-based and paper-pencil tests. *Journal of Educational Computing Research, 31*(1), 51-60. Retrieved on September 9, 2017 from: https://doi-org.iris.etsu.edu:3443/10.2190/GRQQ-YT0F-7LKB-F033

Bond, L. (1996). Norm- and criterion-referenced testing. *Practical Assessment*, 5(2), 1-3. Retrieved on November 10, 2017 from: http://pareonline.net/getvn.asp?v=5%26n=2

Boo, J. & Vispoel, W. (2012) Computer versus paper-and-pencil assessment of educational development: A comparison of psychometric features and examinee preferences. *Psychological Reports, 111*(2), 443-460. Retrieved on November 4, 2017 from: https://doi-org.iris.etsu.edu:3443/10.2466/10.03.11.PR0.111.5.443-460

Brewer, C., Knoeppel, R., & Lindle, J. (2015). Consequential validity of accountability policy: Public understanding of assessments. *Educational Policy, 29*(5), 711-745. DOI: 10.1177/0895904813518099

Buffum, A., Mattos, M., & Weber, C. (2010). The why behind RTI. *Educational Leadership, 68*(2), 10-16. Retrieved on September 9, 2017 from: http://search.ebscohost.com.iris.etsu.edu:2048/login.aspx?direct=true&AuthType=cookie,ip,url,uid,athens&db=eft&AN=508186066&site=ehost-live

Caldwell, D. J., & Pate, A. N. (2013). Effects of question formats on student and item performance. *American Journal of Pharmaceutical Education*, *77*(4), 71. DOI: 10.5688/ajpe77471

Center on Response to Intervention. (2015). Home page. Washington, DC: American Institutes for Research. Retrieved on September 9, 2017 from: http://www.rti4success.org/

Clariana, R. & Wallace, P. (2002). Paper–based versus computer–based assessment: Key factors associated with the test mode effect. *British Journal of Educational Technology*, 33(5), 593-602. DOI: 10.1111/1467-8535.00294

Cizek, G. (2001). More unintended consequences of high-stakes testing. *Educational Measurement: Issues and Practice, 20*(4), 19-27. DOI: 10.1111/j.1745-3992.2001.tb00072.x

Collier, L. (2017). *What does ESSA mean for assessment?* Retrieved on October 2, 2017 from: http://news.ets.org/stories/what-does-essa-mean-for-assessment/

Common Core State Standards (2017). Read the standards. Retrieved on November 19, 2017 from: http://www.corestandards.org/read-the-standards/

Crepeau-Hobson, F, & Biance, M. (2012). Response to intervention. *Intervention in School and Clinic*, *48*(3), 142-151. Retrieved on September 15, 2017 from: https://doi-org.iris.etsu.edu:3443/10.1177/1053451212454005

Darling-Hammond, L., Amrein-Beardsley, A., Haertel, E., & Rothstein, J. (2012). Evaluating teacher evaluation. *Phi Delta Kappan Magazine*, *93*(6), 8-15 DOI: 10.1177/003172171209300603

Deno, S., Reschly, A., Lembke, E., Magnusson, D., Callender, S., Windram, H., . . . Stormont, Melissa. (2009). Developing a school-wide progress-monitoring system. *Psychology in the Schools*, *46*(1), 44-55. Retrieved on September 15, 2017 from: http://dx.doi.org/10.1002/pits.20353

Denton, C. A. (2012). Response to intervention for reading difficulties in the primary grades. *Journal of Learning Disabilities*, *45*(3), 232-243. Retrieved on September 9, 2017 from: https://doi-org.iris.etsu.edu:3443/10.1177/0022219412442155

Dixson, D. and Worrell, F. (2016). Formative and summative assessment in the classroom. *Theory Into Practice, 55*(2), 153-159. DOI: 10.1080/00405841.2016.1148989

Dworkin, A. (2005). The no child left behind act: Accountability, high-stakes testing, and roles for sociologists. *Sociology of Education, 78*(2), 170-174. Retrieved on September 9, 2017 from: https://doi-org.iris.etsu.edu:3443/10.1177/003804070507800205

Educational Testing Services. (2017). Multiple-choice questions — Select one or more answer choices sample questions. Retrieved on September 27, 2017 from: https://www.ets.org/gre/revised_general/prepare/quantitative_reasoning/multiple_choice_more/sample_questions

Edwards, P. A., Turner, J. D., & Mokhtari, K. (2008). Balancing the assessment "of" learning and "for" learning in support of student literacy achievement. *Reading Teacher*, 61(8), 682-684. Retrieved on November 4, 2017 from: https://login.iris.etsu.edu:3443/login?url=https://search-proquest-com.iris.etsu.edu:3443/docview/203280995?accountid=10771

Eklund, K., Renshaw, T., Dowdy, E., Jimerson, S., Hart, S., Jones, C., and Earhart, J. (2009). Early identification of behavioral and emotional problems in youth: Universal screening versus teacher-referral identification. (Report). *The California School Psychologist*, *14*, 89. Retrieved on November 24, 2017 from: https://etsu-edu-primo.hosted.exlibrisgroup.com/primo-explore/fulldisplay?docid=TN_proquest1697788228&context=PC&vid=01ETSU&search_scope=Everything&tab=default_tab&lang=en_US

Faieta, J. (2017). IDEA At age forty: Weathering common core standards and data driven decision making. *Contemporary Issues in Education Research (Online)*, *10*(1), 33-44. Retrieved on September 9, 2017 from: http://dx.doi.org.iris.etsu.edu:2048/10.19030/cier.v10i1.9878

Fletcher, J. M., & Vaughn, S. (2009). Response to intervention: Preventing and remediating academic difficulties. *Child Development Perspectives*, *3*(1), 30–37. DOI**:** 10.1111/j.1750-8606.2008.00072.x

Florida Department of Education. (2015). *Florida standards assessment: Mathematics grade 3.* Florida Department of Education. Tallahassee, FL. Retrieved on September 30, 2017 from: https://fsassessments.org/assets/documents/practice-tests-paper/FSA_3M_Practice-Test_approved.pdf

Gillam, S. & Justice, L. (2010). RTI progress monitoring tools: Assessing primary-grade students in response-to-intervention programs. (response-to-intervention) (Report). *A S H A Leader*, *15*(11), 12-15. Retrieved on September 9, 2017 from: http://go.galegroup.com.iris.etsu.edu:2048/ps/i.do?&id=GALE|A238749623&v=2.1&u=tel_a_etsul&it=r&p=AONE&sw=w&authCount=1

Goldhaber, D. (2015). Exploring the potential of value-added performance measures to affect the quality of the teacher workforce. *Educational Researcher, 44*(2), 87-95. DOI: 10.3102/0013189X15574905

Great Schools Partnership (2014). The glossary of educational reform. Retrieved on November 19, 2017 from: http://edglossary.org

Guskey, Thomas R. (2003). How Classroom Assessments Improve Learning. *Educational Leadership, 60*(5), 6-11. Retrieved on December 9, 2017 from: http://www.ascd.org/publications/educational-leadership/feb03/vol60/num05/How-Classroom-Assessments-Improve-Learning.aspx

Haladyna, T., Haas, N., & Allison, J. (1998). Continuing tensions in standardized testing. *Childhood Education*, *74*(5), 262-273. Retrieved on July 7, 2017 from: http://dx.doi.org/10.1080/00094056.1998.10521950

Hale, J. B. (2008). Response to intervention: Guidelines for parents and practitioners. Wrights law. Retrieved on September 9, 2017 from: http://www.wrightslaw.com/idea/art/rti.hale.pdf

Hallam, R., Lyons, A., Pretti-Frontczak, K., & Grisham-Brown, J. (2014). Comparing apples and oranges. *Topics in Early Childhood Special Education*, 34(2), 106-115. DOI: 10.1177/0271121414524283

Harlen, W., & Deakin-Crick, R. (2002). A systematic review of the impact of summative assessment and tests on students' motivation for learning. In EPPI-Centre (Ed.), Research evidence in education library (1.1 ed., pp. 153–). London, UK: University of London Institute of Education Social Science Research Unit. Retrieved on February 27, 2018 from: http://www.storre.stir.ac.uk/bitstream/1893/19607/1/SysRevImpSummativeAssessment2002.pdf

Herman, J. & Baker, E. (2005). Making Benchmark Testing Work. *Educational Leadership, 63*(3), 48-54. Retrieved on September 30, 2017 from: http://search.ebscohost.com.iris.etsu.edu:2048/login.aspx?direct=true&AuthType=cookie,ip,url,uid,athens&db=eft&AN=507839317&site=ehost-live

Higgins, M. (2009). Standardised tests: Wristwatch or dipstick? *Research in Education*, *81*(1), 1-11. Retrieved on July 7, 2017 from: https://doi-org.iris.etsu.edu:3443/10.7227/RIE.81.1

Jiao, H., & Lissitz, R. W. (2012). Computers and their impact on state assessments. Charlotte, NC: Information Age.

Knight, S., Edmondson, J., Lloyd, G., Arbaugh, F., Nolan, J., Whitney, A., & McDonald, S. (2012). Examining the complexity of assessment and accountability in teacher education. *Journal of Teacher Education*, *63*(5), 301-303. DOI: 10.1177/0022487112460200

Lauen, D., & Gaddis, S. (2012). Shining a light or fumbling in the dark? The effects of NCLB's subgroup-specific accountability on student achievement. *Educational Evaluation and Policy Analysis*, *34*(2), 185-208. Retrieved on September 9, 2017 from: https://doi-org.iris.etsu.edu:3443/10.3102/0162373711429989

Linn, R. L. (2000). Assessments and accountability. *Educational Researcher, 29*(2), 4-16. Retrieved on July 7, 2017 from: https://doi-org.iris.etsu.edu:3443/10.3102/0013189X029002004

Levitt, J. M., Saka, N., Romanelli, L. H., & Hoagwood, K. (2007). Early identification of mental health problems in schools: The status of instrumentation. *Journal of School Psychology*, 45(2), 163-191. Retrieved on November 20, 2017 from: https://doi.org/10.1016/j.jsp.2006.11.005

Lee, J. & Reeves, T. (2012). Revisiting the impact of NCLB high-stakes school accountability, capacity, and resources. *Educational Evaluation and Policy Analysis*,*34*(2), 209-231. Retrieved on September 9, 2017 from: https://doi-org.iris.etsu.edu:3443/10.3102/0162373711431604

Little, M. (2012). Action research and response to intervention: Bridging the discourse divide. *The Educational Forum, 76*(1), 69-80. DOI: 10.1080/00131725.2011.629286

Marzano, R. J. (2010). *Formative assessment and standards based grading: Classroom strategies that work.* Bloomington: Marzano Research Laboratories.

McAllister, D. & Guidice, R. (2012). This is only a test: A machine-graded improvement to the multiple-choice and true-false examination. *Teaching in Higher Education*, 17(2), 193-207. Retrieved on October 28, 2017 from: https://doi.org/10.1080/13562517.2011.611868

McPherson, A. (2016). Criterion-referenced evaluation. *Laboratory* Medicine, 10(5), 306-312. Retrieved on November 4, 2017 from: https://doi.org/10.1093/labmed/10.5.306

McTighe, J. & O'Connor, K. (2005). Seven practices for effective learning. *Educational Leadership*, *63*(3), 10-17. Retrieved on February 27, 2018 from: http://www.ascd.org/publications/educational-leadership/nov05/vol63/num03/Seven-Practices-for-Effective-Learning.aspx

National Center for Learning Disabilities. (2006). *IDEA parent guide: A comprehensive guide to your rights and responsibilities under the Individuals with Disabilities Education Act (IDEA 2004)*. Retrieved on September 9, 2017 from: https://www.ncld.org/wp-content/uploads/2014/11/IDEA-Parent-Guide1.pdf

National Council on Measurement in Education. (2012). Retrieved on September 30, 2017 from: http://www.ncme.org/ncme/NCME/Resource_Center/Glossary/NCME/Resource_Center/Glossary1.aspx?hkey=4bb87415-44dc-4088-9ed9-e8515326a061#anchorI

Nelson, H. (2013). Testing more, teaching less: What America’s obsession with student testing costs in money and lost instructional time. New York: American Federation of Teachers. Retrieved on December 9, 2017 from: https://www.aft.org/sites/default/files/news/testingmore2013.pdf

Nichols, S.L. & Berliner, D.C. (2008). Why has high-stakes testing so easily slipped into contemporary American life? *Phi Delta Kappan, 89*(9), 672-676. Retrieved on December 9, 2017 from: https://doi-org.iris.etsu.edu:3443/10.1177/003172170808900913

Nitko, A. J. and Brookhart, S. M. (2007). *Educational assessment of students* (5th ed.). Upper Saddle Ridge, New Jersey: Pearson Education, Inc.

Pearson. (2009). Computer-based & paper-pencil test comparability studies. *Pearson Assessments*, 1-6. Retrieved on September 30, 2017 from: http://images.pearsonassessments.com/images/tmrs/tmrs_rg/Bulletin_9_rv.pdf?WT.mc_id=TMRS_Bulletin_9_Computer_Based_and_Paper_Pencil

Paulina, A. (c2017). History of standardized testing in the United States. [NEA website]. Retrieved on September 1, 2017 from: http://www.nea.org/home/66139.htm

Popham, W. J. (1999). Why Standardized Test Scores Don't Measure Educational Quality. *Educational Leadership, 56*(6), 8-15. Retrieved on December 9, 2017 from: http://search.ebscohost.com.iris.etsu.edu:2048/login.aspx?direct=true&AuthType=cookie,ip,url,uid,athens&db=eft&AN=507615333&site=ehost-live

Popham, W. J. (2001). Uses and misuses of standardized tests.(educational research). *NASSP Bulletin,84*(622), 24. DOI: 10.1177/019263650108562204

Popham, W. J. (2008). *Classroom assessment: What teachers need to know* (5th ed.). Boston: Pearson Education, Inc.

Pomplun, M., Frey, S., & Becker, D. (2002). The score equivalence of paper-and-pencil and computerized versions of a speeded test of reading comprehension. *Educational and Psychological Measurement*, *62*(2), 337-354. Retrieved on September 15, 2017 from: https://doi-org.iris.etsu.edu:3443/10.1177/0013164402062002009

Renaissance Learning. (2015a). *STAR math technical manual.* Wisconsin Rapids, WI. Retrieved on September 30, 2017 from: https://resources.renlearnrp.com/us/manuals/sm/smrptechnicalmanual.pdf

Renaissance Learning. (2015b). *STAR reading technical manual.* Wisconsin Rapids, WI. Retrieved on September 30, 2017 from: https://resources.renlearnrp.com/us/manuals/sr/srrpsoftwaremanual.pdf

Renaissance EdWords (2017a). *Criterion-referenced test*. Retrieved on November 10, 2017 from: http://www.renaissance.com/2016/09/09/parents-guide-star-assessments/

Renaissance EdWords (2017b). *Norm-referenced test*. Retrieved on November 17, 2017 from: https://www.renaissance.com/edwords/norm-referenced-test/

Renaissance Place (2017). *A parent's guide to STAR assessments*. Retrieved on September 30, 2017 from: https://www.renaissance.com/edwords/criterion-referenced-test/

Risko, V. J., & Walker-Dalhouse, D. (2010). Making the most of assessments to inform instruction. *Reading Teacher, 63*(5), 420-422. DOI: 10.1598/RT/63.5.7

Roderick, M. & Engel, M. (2001). The grasshopper and the ant: Motivational responses of low-achieving students to high-stakes testing. *Educational Evaluation and Policy Analysis, 23*(3), 197-227. Retrieved on December 9, 2017 from: https://doi-org.iris.etsu.edu:3443/10.3102/01623737023003197

RTI Action Network (2017). *What is RTI?* Retrieved August 28, 2017, from The National Center for Learning Disabilities, Inc.: http://www.rtinetwork.org/learn/what/whatisrti

Safer, N. & Fleischman, S. (2005). Research matters: How student progress monitoring improves instruction. *Educational Leadership*, 62(5), p. 81-83. Retrieved on September 9, 2017 from: http://www.ascd.org/publications/educational-leadership/feb05/vol62/num05/How-Student-Progress-Monitoring-Improves-Instruction.aspx

Salinger, R. (2016). Selecting universal screening measures to identify students at risk academically. *Intervention in school and clinic, 52*(2), 77-84. Retrieved on September 30, 2017 from: https://doi-org.iris.etsu.edu:3443/10.1177/1053451216636027

Seo, D. & De Jong, G. (2015). Comparability of online- and paper-based tests in a statewide assessment program. *Journal of Educational Computing Research*, 52(1), 88-113. Retrieved on November 4, 2017 from: https://doi-org.iris.etsu.edu:3443/10.1177/0735633114568856

Shepard, L. (2016). Testing and assessment for the good of education: Contributions of AERA presidents, 1915-2015. *Educational Researcher, 45*(2), 112-121. Retrieved on July 7, 2017 from: https://doi-org.iris.etsu.edu:3443/10.3102/0013189X16639599

Sloane, F., & Kelly, A. (2003). Issues in high-stakes testing programs. *Theory Into Practice, 42*(1), 12-17. DOI: 10.1207/s15430421tip4201_3

Solorzano, R. W. (2008). High stakes testing: issues, implications, and remedies for English language learners. *Review of Educational Research, 78*(2), 260-329. Retrieved on December 9, 2017 from: https://doi-org.iris.etsu.edu:3443/10.3102/0034654308317845

Southern, K. (2015). Assessment in Tennessee: Teaching, testing, and time. *State Collaborative on Reforming Education.* Retrieved on December 9, 2017 from: https://tnscore.org/assessment-in-tennessee-teaching-testing-and-time/

State Collaborative on Reforming Education. (2015) Teaching, testing, and time: Educator voices on improving assessment in Tennessee. Retrieved on December 10, 2017 from: https://view.joomag.com/teaching-testing-and-time/0078588001441993866?short

Stiggins, R. (2002). Assessment crisis: The absence of assessment FOR learning. *Phi Delta Kappan, 83*(10), 758-765. Retrieved on December 9, 2017 from: https://doi.org/10.1177/003172170208301010

Tan, X. & Michel, R. (2001). Why do standardized testing programs report scaled scores? Why not just report the raw or percent-correct scores? *R & D Connections*, 16 (1-6). Retrieved on August 5, 2017 from: https://www.ets.org/Media/Research/pdf/RD_Connections16.pdf

Thorius, K., & Maxcy, B. (2015). Critical practice analysis of special education policy. *Remedial and special education,36*(2), 116-124. Retrieved on September 9, 2017 from: https://doi-org.iris.etsu.edu:3443/10.1177/0741932514550812

Tienken, C. H. (2015). Test use and abuse. *Kappa Delta Pi Record, 51*(4), 155-158. Retrieved on December 10, 2017 from: http://dx.doi.org/10.1080/00228958.2015.1089617

TEAM (n.d.a). *Data: Achievement measures*. Retrieved December 10, 2017 from Tennessee Department of Education: http://team-tn.org/data/achievement-measures/

TEAM (n.d.b). *Data: Growth measures*. Retrieved December 10, 2017 from Tennessee Department of Education: http://team-tn.org/data/growth-measures/

TEAM (n.d.c). *Data: TVAAS*. Retrieved December 10, 2017 from Tennessee Department of Education: http://team-tn.org/data/tvaas/

TEAM (n.d.d). *Evaluation: Teacher evaluation*. Retrieved December 10, 2017 from Tennessee Department of Education: http://team-tn.org/evaluation/teacher-evaluation/

Tennessee State Board of Education (2017). *Teacher and administrator evaluation policy*. Retrieved on December 10, 2017 from: https://www.tn.gov/content/dam/tn/stateboardofeducation/documents/5.201_Teacher__Principal_Evaluation_Policy_7-28-17.pdf

Tennessee Department of Education (2015). *Guide to test report interpretation*. Retrieved November 10, 2017 from Tennessee Department of Education. Interpretation Guide is updated every year: https://www.tn.gov/content/dam/tn/education/testing/tnready/tnready_guide_test_interpretation_eoc_2017-18.pdf

Tennessee Department of Education (2017a). *Every student succeeds act: Building on success in Tennessee*. Retrieved September 22, 2017 from Tennessee Department of Education: https://www.tn.gov/education/essa.html

Tennessee Department of Education (2017b). *2017-18 TNReady item types*. Retrieved September 20, 2017 from Tennessee Department of Education: https://www.tn.gov/content/dam/tn/education/documents/tnready_item_types_2017-18.pdf

Tennessee Department of Education (n.d.a). *2017-2018 Assessment FAQ*. Retrieved November 18, 2017, from Tennessee Department of Education: https://www.tn.gov/education/assessment/tnready/assessment-faq.html

Tennessee Department of Education (n.d.b). *How is TNready created for Tennessee students?* Retrieved September 22, 2017, from Tennessee Department of Education: https://www.tn.gov/content/dam/tn/education/documents/How_TNReady_is_Created_FINAL.pdf

Tennessee Department of Education (n.d.c). *Overview of testing in Tennessee*. Retrieved August 28, 2017, from Tennessee Department of Education: https://www.tn.gov/education/assessment/testing-overview.html

Tennessee Department of Education (n.d.d). *Parent guide to being TNReady: Preparing for the 2016-17 school year*. Retrieved September 22, 2017 from Tennessee Department of Education. Parent guide is updated every year: https://www.tn.gov/content/dam/tn/education/documents/TNReady_Parent_Guide_2017-18.pdf

Tennessee Department of Education (n.d.e). *Student assessment in Tennessee*. Retrieved September 20, 2017, from Tennessee Department of Education: https://www.tn.gov/education/assessment.html

Tennessee Department of Education (n.d.f). *Testing times by grade/subject*. Retrieved December 9, 2017, from Tennessee Department of Education. Test times are updated every year: https://www.tn.gov/education/assessment/testing-times-by-grade-subject.html

Tennessee Department of Education (n.d.g). *TNReady*. Retrieved September 24, 2017 from Tennessee Department of Education: https://www.tn.gov/education/assessment/tnready.html

Tennessee Department of Education (n.d.h). *TNReady grades 3-5 mathematics 2017-2018 school year*. Retrieved November 4, 2017, from Tennessee Department of Education: https://www.tn.gov/content/dam/tn/education/documents/tnready_blueprints_math_grade_3-5.pdf

Tomlinson, C. A., & Moon, T. R. (2013). *Assessment and student success in a differentiated classroom*. ASCD. Retrieved on February 27, 2018 from: http://www.ascd.org/ASCD/pdf/siteASCD/publications/assessment-and-di-whitepaper.pdf

United States. (1992). *Testing in American schools: Asking the right questions*. Congress of the United States, Office of Technology Assessment. Retrieved on July 7, 2017.

United States Department of Education. (n.d.) *Every Student Succeeds Act (ESSA)*. Retrieved on October 2, 2017 from: https://www.ed.gov/esea

United States Department of Education. (2015) *Fact sheet: Testing action plan*. Retrieved on December 9, 2017 from: https://www.ed.gov/news/press-releases/fact-sheet-testing-action-plan

Van de Watering, G., Gijbels, D., Dochy, F., & Van der Rijt, J. (2008). Students' assessment preferences, perceptions of assessment and their relationships to study results. *Higher Education, 56*(6), 645-658. DOI**:** 10.1007/s10734-008-9116-6

Van Horn, R. (2003). Computer adaptive tests and computer-based tests. *Phi Delta Kappan Magazine, 84*(8), 567-631. Retrieved on September 9, 2017 from: https://doi-org.iris.etsu.edu:3443/10.1177/003172170308400804

Vanderheyden, A. (2013). Universal screening may not be for everyone: Using a threshold model as a smarter way to determine risk. School psychology review. 42. 402-414. Retrieved on September 9, 2017 from: https://www.researchgate.net/publication/281207270_Universal_Screening_May_Not_Be_for_Everyone_Using_a_Threshold_Model_as_a_Smarter_Way_to_Determine_Risk

Vispoel, W. P., Boo, J., & Bleiler, T. (2001). Computerized and paper-and-pencil versions of the Rosenberg Self-Esteem Scale: a comparison of psychometric features and respondent preferences. *Educational and Psychological Measurement, 61*(3), 461- 474. Retrieved on November 4, 2017 from: https://doi-org.iris.etsu.edu:3443/10.1177/00131640121971329

Wang, S., Jiao, H., Young, M. J., Brooks, T., & Olson, J. (2008). Comparability of Computer-Based and Paper-and-Pencil Testing in K-12 Reading Assessments: A Meta-Analysis of Testing Mode Effects. Educational and Psychological Measurement, 68(1), 5-24. Retrieved on November 4, 2017 from: https://doi-org.iris.etsu.edu:3443/10.1177/0013164407305592

Weiss, David J. (2004). Computerized adaptive testing for effective and efficient measurement in counseling and education. *Measurement and Evaluation in Counseling and Development, 37*(2), 70-84. Retrieved on September 15, 2017 from: http://iacat.org/sites/default/files/biblio/we04070.pdf

Wiliam, Dylan. (2011). What Is Assessment for Learning? *Studies in Educational Evaluation,37*(1), 3-14. DOI**:** 10.1016/j.stueduc.2011.03.001

Wiliam, D. (2017). Assessment: The bridge between teaching and learning. *ATA Magazine, 97*(4), 28-32. Retrieved on February 27, 2018 from https://search-proquest-com.iris.etsu.edu:3443/docview/1924811458?accountid=10771&rfr_id=info%3Axri%2Fsid%3Aprimo

Wixson, K. K., & Valencia, S. W. (2011). Assessment in RTI: What teachers and specialists need to know. *The Reading Teacher, 64*(6), 466-469. Retrieved on September 15, 2017 from: https://login.iris.etsu.edu:3443/login?url=https://search-proquest-com.iris.etsu.edu:3443/docview/857079370?accountid=10771

Zucker, S. (2004). Administration practices for standardized assessment. Pearson Education, Inc. Retrieved on November 18, 2017 from: https://images.pearsonassessments.com/images/tmrs/tmrs_rg/AdministrationPractices.pdf?WT.mc_id=TMRS_Administration_Practices_for_Standardized_Assessments

APPENDICES

APPENDIX A

**Types of Questions on Educational Assessments**

Example of a Multiple-Choice Question (Nitko & Brookhart, 2007):

In what year did the United States enter World War I?

A. 1776
B. 1812
C. 1917
D. 1981

Example of a Multiple-Choice Question (Florida Department of Education, 2015):

What is the value of the unknown number in the equation 6 x 3 = ☐ ?

A. 13
B. 19
C. 18
D. 63

Example of a Matching Exercise (Marzano, 2010):

Match the state listed on the left with its most famous landmark listed on the right.

| | |
|---|---|
| 1. District of Columbia _____ | 5. The Alamo |
| 2. Arizona _____ | 6. The Pentagon |
| 3. South Dakota _____ | 7. The White House |
| 4. Texas _____ | 8. Mount Rushmore |
| | 9. Everglades |
| | 10. Grand Canyon |

Example of a matching table sample questions from Florida Department of Education (2015):

Match each number to the value of the number rounded to the nearest 10.

| | 180 | 190 | 200 |
|---|---|---|---|
| 181 | Ⓐ | Ⓑ | Ⓒ |
| 186 | Ⓓ | Ⓔ | Ⓕ |
| 194 | Ⓖ | Ⓗ | Ⓘ |

Example of an Alternative-Choice Question (Marzano, 2010):

"Sally sold seashells" is an example of:

A. Hyperbole
B. Alliteration

Example of a true/false item:

____Quotation marks are used at the end of statements that are questions.

____ The intersection of any two lines produces four 45-degree angles.

Example of a multiple response item (Marzano, 2010):

Put a check next to shapes for which you can find the volume:

_____Circle

_____Cube

_____Square

_____Sphere

_____Octagon

_____Prism

Example from Educational Testing Services (ETS) (2107):

Which of the following integers are multiples of both 2 and 3?

Indicate <u>all</u> such integers.

A. 8
B. 9
C. 12
D. 18
E. 21
F. 36

Example questions from Florida Department of Education (2015):

Select all the fractions that are equivalent to a whole number.

Ⓐ $\frac{3}{3}$

Ⓑ $\frac{5}{10}$

Ⓒ $\frac{8}{2}$

Ⓓ $\frac{15}{7}$

Ⓔ $\frac{1}{6}$

Select all the expressions that are equal to 324.

A. 372 – 48
B. 660 – 346
C. 119 + 215

D. 728 – 404

E. 216 + 108

Example of a fill in the blank item (Marzano, 2010):

1. A fraction in which the numerator is greater than the denominator is a(n) ________ fraction.
2. Animals that only eat vegetation are called _____, animals that eat only meat are called _____, and animals that eat both vegetation and meat are called_____.

Example of a short answer item (Nitko & Brookhart, 2007)"

1. What is the pen name of the author of *Alice in Wonderland*? ___________________
2. What city is located at the confluence of the Allegheny and Monongahela rivers? ___________________________

Examples of a short answer item (Parent guide to being TNReady: Preparing for the 2016-17, TNDOE, n.d.d):

1. Makenna purchases a car for $27,500. The value of the car will depreciate each year. After five years, the value of the car is $14,186. What is the approximate yearly depreciation rate of the car, to the nearest tenth of a percent?

2. Evaluate $39 - (11 + 5^3 \div 5)$

APPENDIX B

## Permission Letters

### Permission Letter for Administrator Participation in the Collection of Data

January 11, 2018
Dr. Jason Vance
Director of Schools
Loudon County School System
100 River Road
Loudon, Tennessee 37774
865-458-5411

Dear Dr. Vance:

I am working to complete my Doctoral Dissertation at East Tennessee State University, Johnson City, TN. I would like to gain your permission to request participation from administrators within your system. Per our previous conversation, the collection of the data will be performed by the participating administrators to protect student privacy as well as teacher privacy. Participating administrators will provide the researcher with a data set of numerical test scores only, no identifiable information from students will be included.

If these arrangements meet with your approval, please sign the letter where indicated below and return it to me in the enclosed return envelope.

Thank you very much.

Sincerely,

Brooke Sampson

**PERMISSION GRANTED:**

I have the authority to grant the permission requested herein, and I hereby grant Brooke Sampson permission to request participation from administrators within the district under the above listed guidelines in pursuit of her doctoral study.

Jason Vance

Signature

Jason Vance

Name

100 River Rd

Loudon, TN 37774

Address

1/12/2008

Date

**Permission Letter for Administrator Participation in the Collection of Data**

January 9, 2018
Debra Ann Cline, Ed.D
Assistant Superintendent
Sevier County School System
226 Cedar Street
Sevierville, Tennessee 37862
865-453-4671

Dear Dr. Cline:

I am working to complete my Doctoral Dissertation at East Tennessee State University, Johnson City, TN. I would like to gain your permission to request participation from administrators within your system. Per our previous conversation, the collection of the data will be performed by the participating administrators to protect student privacy as well as teacher privacy. Participating administrators will provide the researcher with a data set of numerical test scores only, no identifiable information from students will be included.

If these arrangements meet with your approval, please sign the letter where indicated below and return it to me in the enclosed return envelope.

Thank you very much.

Sincerely,

Brooke Sampson

**PERMISSION GRANTED:**

I have the authority to grant the permission requested herein, and I hereby grant Brooke Sampson permission to request participation from administrators within the district under the above listed guidelines in pursuit of her doctoral study.

* Debra Ann Cline

Signature

Debra A. Cline, Assistant Superintendent

Name

226 Cedar Street

Sevierville, Tennessee 37862

Address

Date 1/16/18

* Participation by principals will be on a voluntary basis.

APPENDIX C

## STAR Math Instructional Planning Report

STAR Math

# Instructional Planning Report for [redacted]

Printed [redacted]



School: [redacted]
Class: [redacted]

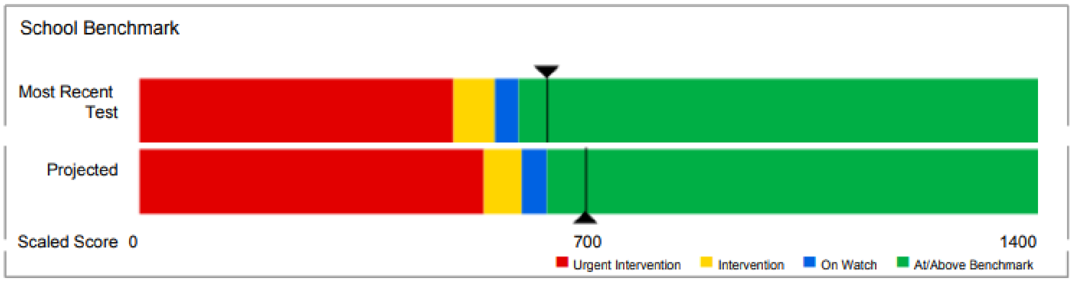
Teacher: [redacted]
Grade: [redacted]

**Report Options**
Use Trend Score: Use trend score for student's suggested skills

### STAR Math Enterprise Test Results

Current SS (Scaled Score): 643 | Test Date: 12/13/2017[a] | Audio enabled for this test.
Projected SS for 06/01/18: 704 | Based on research, 50% of students at this student's level will achieve this much growth.

### [redacted] Current Performance




### Suggested Skills

[redacted] STAR Math scaled score(s) suggest these skills from Core Progress Math built for TN learning progressions would be challenging, but not too difficult for her. Combine this information with your own knowledge of the student and use your professional judgment when designing an instructional program. Use the Core Progress Math built for TN learning progressions to see how these skills fit within the larger context of the progression.

**Geometry**

| GR | |
|---|---|
| | **Geometry**<br>This score suggests [redacted] is ready for instruction and practice with the following skills. |
| 4 | » Draw a line of symmetry |

**Measurement and Data**

| GR | |
|---|---|
| | **Measurement and Data**<br>This score suggests [redacted] is ready for instruction and practice with the following skills. |
| 4 | Solve an addition or subtraction problem to find an unknown angle measure on a diagram |
| 4 | » Solve a problem using the area and perimeter formulas for rectangles |
| 4 | Represent measurement data with a line plot that has a fractional scale to eighths |
| 4 | Solve a problem involving addition or subtraction of fractions by using information presented in a line plot with a fractional scale to eighths |

» Designates a focus skill. Focus skills identify the most critical skills to learn at each grade level.

[a] This student was given additional time to complete the assessment.

STAR™ Math

# Instructional Planning Report for [redacted]

Printed [redacted]

School: [redacted] Teacher: [redacted]
Class: [redacted] Grade: [redacted]

## Numbers and Operations

| GR | Number and Operations in Base Ten |
|---|---|
| | This score suggests [redacted] is ready for instruction and practice with the following skills. |
| 5 | » Fluently multiply multi-digit whole numbers |
| 5 | » Divide a whole number of up to four digits by a 2-digit whole number using one of various strategies |
| 5 | Demonstrate the reasoning used in a division problem with a dividend of up to four digits and a 2-digit divisor |

| GR | Number and Operations - Fractions |
|---|---|
| | This score suggests [redacted] is ready for instruction and practice with the following skills. |
| 5 | Add or subtract fractions or mixed numbers with unlike denominators |
| 5 | Estimate a sum or difference of two fractions |
| 5 | Assess the reasonableness of an answer to an addition or subtraction problem involving fractions |
| 5 | » Solve problems involving addition or subtraction of fractions referring to the same whole |
| 5 | Interpret a fraction as division of the numerator by the denominator |
| 5 | Divide whole numbers to solve a problem, leading to an answer in the form of a fraction or a mixed number |
| 5 | Interpret the product of a fraction and a fraction or whole number as parts of a partition or a sequence of operations |
| 5 | Relate a fraction product to a rectangular area model |
| 5 | Model multiplication of fractions by finding the area of a rectangle using tiling |
| 5 | » Multiply a fraction by a fraction |
| 5 | » Multiply fractional side lengths of a rectangle to find its area |
| 5 | Compare the size of a product of two fractions to the size of one factor based on the size of the other factor |

» Designates a focus skill. Focus skills identify the most critical skills to learn at each grade level.

ª This student was given additional time to complete the assessment.

APPENDIX D

## STAR Reading Instructional Report

STAR™ Reading

# Instructional Planning Report

# for [redacted]

Printed [redacted]

1

School: [redacted]
Teacher: [redacted]
Class: [redacted]
Grade: [redacted]

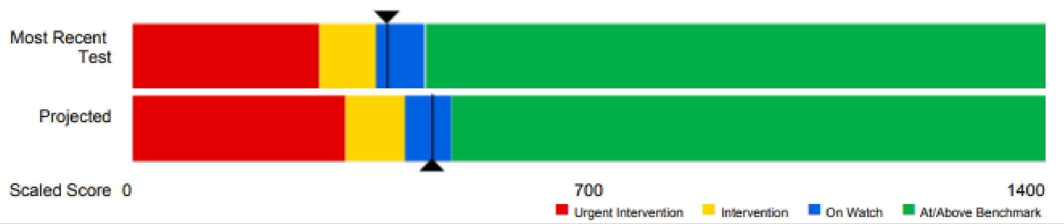
**Report Options**
Use Trend Score: Use trend score for student's suggested skills

### STAR Reading Test Results

| | |
|---|---|
| Current SS (Scaled Score): 395 | Test Date: 12/08/2017[a] |
| IRL: 3.4 | ZPD: 2.8-4.1 |
| Projected SS for 06/01/18: 465 | Based on research, 50% of students at this student's level will achieve this much growth. |

### [redacted] Current Performance

School Benchmark




### Suggested Skills

[redacted] STAR Reading scaled score(s) suggest these skills from Core Progress Reading built for TN learning progressions would be challenging, but not too difficult for her. Combine this information with your own knowledge of the student and use your professional judgment when designing an instructional program. Use the Core Progress Reading built for TN learning progressions to see how these skills fit within the larger context of the progression.

| GR | Reading Literature |
|---|---|
| | **Craft and Structure** |
| | This score suggests [redacted] is ready for instruction and practice with the following skills. |
| 4 | » Explain the meanings of simple similes, metaphors, and uses of exaggeration (e.g., as bright as the sun, a ton of homework) in grade-appropriate texts |

| GR | |
|---|---|
| | **Range of Reading and Level of Text Complexity** |
| | This score suggests [redacted] is ready for instruction and practice with the following skills. |
| 4 | » Read regularly and independently in fourth-grade-appropriate texts for sustained periods of time, increasing speed, stamina, and comprehension |

| GR | Reading Informational Text |
|---|---|
| | **Key Ideas and Details** |
| | This score suggests [redacted] is ready for instruction and practice with the following skills. |
| 5 | Confirm or refute predictions and make adjustments by weighing information against prior experience and knowledge of the text |

» Designates a focus skill. Focus skills identify the most critical skills to learn at each grade level.

[a] This student was given additional time to complete the assessment.

STAR™ Reading

# Instructional Planning Report
## for [redacted]

Printed [redacted]

School: [redacted] Teacher: [redacted]
Class: [redacted] Grade: [redacted]

| GR | **Craft and Structure** |
|---|---|
| | This score suggests [redacted] is ready for instruction and practice with the following skills. |
| 5 | Identify and analyze common organizational structures to determine the connections between ideas (e.g., cause/effect, main idea/support) |

| GR | **Integration of Knowledge and Ideas** |
|---|---|
| | This score suggests [redacted] is ready for instruction and practice with the following skills. |
| 5 | Distinguish facts from opinions, and recognize that opinions and inferences can be supported by facts |

| GR | **Range of Reading and Level of Text Complexity** |
|---|---|
| | This score suggests [redacted] is ready for instruction and practice with the following skills. |
| 4 | » Read regularly and independently in fourth-grade-appropriate texts for sustained periods of time, increasing speed, stamina, and comprehension |

### Foundational Literacy

| GR | **Phonics and Word Recognition** |
|---|---|
| | This score suggests [redacted] is ready for instruction and practice with the following skills. |
| 5 | Use knowledge of syllabication patterns to accurately read unfamiliar multisyllabic words in context (e.g., read *multiple* in a contextual sentence) |
| 5 | Accurately read unfamiliar grade-appropriate multisyllabic words in context using knowledge of all letter-sound correspondences (e.g., the ice on the pond was *solid*) |
| 5 | » Use knowledge of syllabication patterns to accurately read unfamiliar multisyllabic words in isolation (e.g., election, cabinet) |
| 5 | » Accurately read unfamiliar grade-appropriate multisyllabic words in isolation using knowledge of all letter-sound correspondences (e.g., solid, enamel) |
| 5 | Accurately read unfamiliar grade-appropriate multisyllabic words in context, applying greater knowledge of word morphology (e.g., read *visible* in a contextual sentence) |
| 5 | » Accurately read unfamiliar grade-appropriate multisyllabic words in isolation, applying greater knowledge of word morphology (e.g., destruction, transport) |

| GR | **Vocabulary Acquisition** |
|---|---|
| | This score suggests [redacted] is ready for instruction and practice with the following skills. |
| 4 | » Relate grade-appropriate words to their synonyms (e.g., tale/story/myth, fearful/afraid/terrified) and describe how they differ slightly in meaning |
| 4 | » Use knowledge of grade-appropriate affixes (e.g., dis-, in-, mis-, -ion, -less, -ment), base words (e.g., arm, fear), and Greek and Latin roots (e.g., act, graph) in order to predict the meanings of unfamiliar or complex words (e.g., fearless, biography) |
| 4 | Use knowledge of word relationships (e.g., similarities, contrasts) to clarify the meanings of words in a text |
| 4 | Use context clues (e.g., synonyms, antonyms, examples, definitions, restatement) to determine or clarify the meanings of general academic and content-area words or phrases |

» Designates a focus skill. Focus skills identify the most critical skills to learn at each grade level.

ª This student was given additional time to complete the assessment.

STAR™ Reading

# Instructional Planning Report

## for [redacted]

Printed [redacted]



School: [redacted] Teacher: [redacted]

Class: [redacted] Grade: [redacted]

| GR | |
|---|---|
| | **Vocabulary Acquisition**<br>This score suggests [redacted] is ready for instruction and practice with the following skills. |
| 4 | » Use grade-appropriate general academic vocabulary, including words that indicate precise actions, emotions, and states of being (e.g., whined, loneliness, peacefulness) as well as grade-appropriate content-area vocabulary (e.g., wildlife, adapt, habitat) correctly in context |
| 4 | » Use the correct homophones (e.g., weave/we've) and homographs/multiple-meaning words (e.g., sentence, crowd) and determine their meanings in grade-appropriate texts using sentence context or prior knowledge of spellings |
| 4 | Recognize and explain the meanings of common idioms and adages (e.g., once in a blue moon) |
| 4 | » Explain the meanings of simple similes, metaphors, and uses of exaggeration (e.g., as bright as the sun, a ton of homework) in grade-appropriate texts |
| 4 | Recognize that words with similar meanings (e.g., house and home) evoke different responses |
| 4 | Recognize when the context in which a word is used impacts its meaning (e.g., *nice* when used sincerely or when used ironically) |

» Designates a focus skill. Focus skills identify the most critical skills to learn at each grade level.

[a] This student was given additional time to complete the assessment.

# VITA

## BROOKE A. SAMPSON

| | |
|---|---|
| Personal Data: | Date of Birth: May 17, 1983 |
| | Place of Birth: Knoxville, Tennessee |
| | Marital Status: Married |
| Education: | Public Schools, Sevierville, Tennessee |
| | A.A.S. Elementary Education, Walters State Community College, Morristown, Tennessee, 2006 |
| | B.S. Elementary Education, East Tennessee State University, Johnson City, Tennessee, 2008 |
| | M.A. Curriculum and Instruction, East Tennessee State University, Johnson City, Tennessee, 2009 |
| | Ed.S. Educational Administration and Supervision, Lincoln Memorial, Harrogate, Tennessee, 2010 |
| | Ed.D. Educational Leadership, East Tennessee State University, Johnson City, Tennessee, 2018 |
| Professional Experience: | Teacher, Seymour Intermediate School; Seymour Tennessee, 2008-2012 |
| | Teacher, Sevierville Intermediate School; Sevierville, Tennessee, 2012- present |