

12-2016

Effects of TLI- Versus NWEA-Interim Assessment on Academic Achievement of Sixth- and Seventh-Grade Students

Walt Davis
Harding University

Follow this and additional works at: <https://scholarworks.harding.edu/hu-etd>

 Part of the [Educational Assessment, Evaluation, and Research Commons](#)

Recommended Citation

Davis, Walt, "Effects of TLI- Versus NWEA-Interim Assessment on Academic Achievement of Sixth- and Seventh-Grade Students" (2016). *Dissertations*. 7.

<https://scholarworks.harding.edu/hu-etd/7>

This Dissertation is brought to you for free and open access by Scholar Works at Harding. It has been accepted for inclusion in Dissertations by an authorized administrator of Scholar Works at Harding. For more information, please contact scholarworks@harding.edu.



EFFECTS OF TLI- VERSES NWEA-INTERIM ASSESSMENT ON ACADEMIC
ACHIEVEMENT OF SIXTH- AND SEVENTH-GRADE STUDENTS

by

Walt Davis

Dissertation

Submitted to the Faculty of

Harding University

Cannon-Clary College of Education

in Partial Fulfillment of the Requirements for

the Degree of

Doctor of Education

in

Educational Leadership P-20

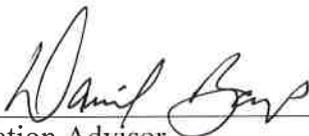
December 2016

EFFECTS OF TLI- VERSUS NWEA-INTERIM ASSESSMENT ON ACADEMIC
ACHIEVEMENT OF SIXTH- AND SEVENTH-GRADE STUDENTS

by

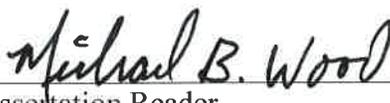
Walt Davis

Dissertation



Dissertation Advisor

9-20-16
Date



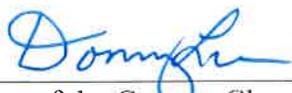
Dissertation Reader

9-20-16
Date



Dissertation Reader

9-20-16
Date



Dean of the Cannon-Clary College of Education

9-20-16
Date



Provost

10/12/16
Date

©2016

Walt Davis

All Rights Reserved

ACKNOWLEDGMENTS

I would like to thank the following people whose encouragement and support guided through my studies and the completion of this dissertation:

To Dr. David Bangs, committee chairperson, for the confidence he displayed in me as a candidate and a person. Because of his guidance and encouragement both professionally and personally, I knew that I would complete this study.

To Dr. Clara Carroll, committee member, for her suggestions that were helpful.

To Dr. Mike Wood, committee member, for giving his time and expertise.

To Drs. Usen Akpanudo and Michael Brooks, for their shared expertise and encouragement that made completion a possibility.

To my daughters, Ticia and Callie, for their love, which always helped me continue to push towards completion.

To my sister, Susan Reynolds, a life-long educator, who always has supported and loved me through thick and thin.

DEDICATION

This dissertation is dedicated to my wife, Tina Davis, and my parents, Robert (Sonny) and Emily Davis. Tina has supported me through the dissertation and has adapted as the wife of a life-long teacher, coach, administrator and learner. Her love inspires me. Mom and Dad, both life-long educators, have been inspirations through example. They have always given me unconditional support and love in education and life.

ABSTRACT

by
Walt Davis
Harding University
December 2016

Title: Effects of TLI- Versus NWEA-Interim Assessment on Academic Achievement of Sixth- and Seventh-Grade Students (Under the direction of Dr. David Bangs)

This project focused on two online providers, The Learning Institute (TLI) and Northwest Evaluation Association (NWEA). The purpose of this dissertation was to provide additional research concerning the effects of interim assessments (TLI versus NWEA) and gender on literacy and mathematics achievement for students in Grades 6 and 7 in Northwest Arkansas. Of the four included school districts, two used TLI and two used NWEA as their interim assessment. For all four hypotheses, the independent variables were gender and interim assessment type. In the first and second hypotheses, the dependent variable was literacy achievement as measured by student performance on the 2013 Arkansas Benchmark sixth- and seventh-grade literacy examinations, respectively. Similarly, in the third and fourth hypothesis, the dependent variable was mathematics achievement as measured by student performance on the 2013 Arkansas Benchmark sixth- and seventh-grade mathematics examinations. Through a review of the literature, the various aspects of the effects of gender and interim assessment type on student achievement in reading and mathematics were identified.

The sample for this study included students from two school districts who used TLI as an interim assessment and two school districts who used NWEA as an interim assessment. The first TLI district had 306 sixth-grade students and 321 seventh-grade students. Of the sixth-grade students, 51% were females, and 63% of the students were Caucasian. Of the seventh-grade students, 50% were females and 67% of the students were Caucasian. The second TLI district had 1,135 sixth-grade students and 1,010 seventh-grade students. Of the sixth-grade students, 50% were females, and 54% of the students were Caucasian. Of the seventh-grade students, 50% were females, and 56% of the students were Caucasian. The first NWEA school district has 680 sixth-grade students and 691 seventh-grade students. Of the sixth-grade students, 47% were females, and 69% of the students were Caucasian. Of the seventh-grade students, 49% were females, and 70% of the students were Caucasian. The second NWEA school district had 1,509 sixth-grade students and 1,501 seventh-grade students. Of the sixth-grade students, 48% were females, and 40% of the students were Caucasian. Of the seventh-grade students, 48% were females, and 39% of the students were Caucasian. All of the selected school districts were within a 50 mile radius of one another.

A 2 x 2 factorial ANOVA was used to analyze the data collected for each of the four hypotheses. The results of this study indicated no significant interaction effects between gender and interim assessment type on literacy and mathematics achievement measured by Arkansas Benchmark Literacy and Mathematics Exams for Grades 6 and 7. In all four hypotheses, no significant interaction effect existed. For Hypotheses 1 and 2, the main effect of gender and assessment was statistically significant. Students using the TLI assessment performed better on the Benchmark Literacy Exam compared to students

using the NWEA assessment regardless of their gender. Additionally, females outperformed males on the Benchmark Literacy Exam for sixth-grade students regardless of the type of assessment. However, for Hypotheses 3 and 4, the main effect of gender and assessment were not significantly significant. Further, there was no significant difference in mathematics achievement by gender or by assessment type.

While some studies were found that showed the effect of gender or interim assessment on academic achievement, no studies were found that compared two interim assessment programs. The results from the studies were mixed. Some research identified significant difference in academic achievement while others did not. Most research deemed interim assessment to be beneficial but was inconsistent as to what degree. This study provided new research comparing competitive interim assessment programs. TLI students performed better on the Benchmark Literacy Exam than NWEA students for both grades examined. Also, females outperformed males on the Benchmark Literacy Exam for both grades examined. The review of mathematics performance for both grades revealed no significant difference by gender or assessment. The study provided data for school leaders examining the effect of interim assessments.

TABLE OF CONTENTS

LIST OF TABLES	xi
CHAPTER I—INTRODUCTION.....	1
Statement of the Problem.....	3
Background.....	3
Hypotheses	16
Description of Terms.....	17
Significance	19
Process to Accomplish	20
CHAPTER II—REVIEW OF RELATED LITERATURE	24
Online Interim Assessment Programs	25
Assessment and Curriculum Alignment.....	32
Gender	35
Conclusion	39
CHAPTER III—METHODOLOGY.....	41
Research Design.....	43
Sample.....	44
Demographics.....	44
Instrumentation.....	45
Data Collection Procedures	46

Analytical Methods	47
Limitations	47
CHAPTER IV—RESULTS	50
Hypothesis 1	50
Hypothesis 2	53
Hypothesis 3	57
Hypothesis 4	60
Summary	64
CHAPTER V—DISCUSSION	66
Conclusions	67
Implications	71
Recommendations	74
REFERENCES	80

LIST OF TABLES

1. Descriptive Statistics from Sixth-Grade 2013 Arkansas Benchmark Literacy Exam Raw Scores	51
2. Factorial ANOVA Results from 2013 Arkansas Sixth-Grade Benchmark Literacy Exam Raw Scores	52
3. Descriptive Statistics from Seventh-Grade 2013 Arkansas Benchmark Literacy Exam Raw Scores	54
4. Factorial ANOVA Results from 2013 Arkansas Seventh-Grade Benchmark Literacy Exam Raw Scores	55
5. Descriptive Statistics from Sixth-Grade 2013 Arkansas Benchmark Mathematics Exam Raw Scores.....	58
6. Factorial ANOVA Results from 2013 Arkansas Sixth-Grade Benchmark Mathematics Exam Raw Scores.....	59
7. Descriptive Statistics from Seventh-Grade 2013 Arkansas Benchmark Mathematics Exam Raw Scores.....	61
8. Factorial ANOVA Results from 2013 Arkansas Seventh-Grade Benchmark Mathematics Exam Raw Scores.....	62
9. Summary of Statistically Significant Results for Hypotheses 1-4.....	64

CHAPTER I

INTRODUCTION

A recent trend in education is for school leaders to seek support from web-based companies to help districts with interim and formative assessments for improvement of student performance. These services can also benefit the alignment of curriculum. Axelson (2005) noted that any online assessment is attractive to educators because student data may be gathered in large amounts and accessed quickly. However, the author noted that districts should exercise caution when purchasing such web-based assessments. Many vendors oversell the correlation between their assessment items and the state standards, causing some states to be hesitant to recommend any vendors. However, many school districts attempt to take advantage of these types of services. This project focused on two online providers: The Learning Institute (TLI) and the Northwest Evaluation Association (NWEA).

Three important forms of assessment are summative, interim, and formative. Perie, Marion, and Gong (2007) said, “Summative assessments are generally given one time at the end of some unit of time such as the semester or school year to evaluate students’ performance against a defined set of standards” (p. 1). State required end-of-course and state benchmark exams are examples of summative assessments. Southworth (2008) said, “Formative assessments are part of the instruction designed to provide crucial feedback for teachers and students. Assessment results inform the teacher of what

has been taught well and not so well. They inform students of what they learn well...” (p. 1). Crooks (2001) stated, “Summative assessment is intended to summarize student attainment at a particular time, whereas formative assessment is intended to promote further improvement of student attainment” (p. 1). Effective formative assessment may positively impact students’ summative assessment performance. The origin of the terms *formative* and *summative assessment* dates back to the late 1960s. Michael Slavin coined the terms, *formative* and *summative assessment*, and Benjamin Bloom made formative assessments building blocks of his Learning for Mastery (Southworth, 2008). Curriculum mapping has a more modern origin than formative assessment. According to their *Research Brief*, Appalachia Educational Laboratory (2005) noted, “Webb (1997) defines alignment as the degree to which expectations [standards] and assessments are in agreement and serve in conjunction with one another to guide the system towards students learning what they are expected to know and do” (p. 1). Formative assessments, interim assessments, and curriculum alignment are tools provided by TLI and NWEA to assist instruction.

Consequently, interim assessment fits between formative and summative. Perie et al. (2007) noted that interim assessment evaluates students’ knowledge and performance related to a specific set of goals over a limited period. Additionally, interim assessments provide data for decisions within the classroom but are also used at a broader level including predicting success for a high stakes summative assessment. Interim assessments are valuable in identifying what students have learned during sections of the course. Summative, interim, and formative assessments each provide unique data to guide instructors and administrators into the best possible decisions regarding student success.

Statement of the Problem

The purpose of this study is four-fold. First, the purpose of this study is to determine by gender the results of TLI-interim assessment versus NWEA-interim assessment on Benchmark Literacy Exam performance for sixth-grade students from Arkansas middle schools. Second, the purpose of this study is to determine by gender the results of TLI-interim assessment versus NWEA-interim assessment on Benchmark Literacy Exam performance for seventh-grade students from Arkansas middle schools. Third, the purpose of this study is to determine by gender the results of TLI-interim assessment versus NWEA-interim assessment on Benchmark Mathematics Exam performance for sixth-grade students from Arkansas middle schools. Fourth, the purpose of the study is to determine by gender the results of TLI-interim assessment versus NWEA-interim assessment on Benchmark Mathematics Exam performance for seventh-grade students from Arkansas middle schools.

Background

Two general vendors that provide assessment and alignment opportunities are TLI and NWEA, which are frequently used in Arkansas public schools. They were compared and contrasted in several ways throughout this dissertation project. Both of these online providers use various assessments and curriculum alignment to inform administrators and teachers with data. The goal is that the information help educators improve teacher instruction and lead to improved student success. Formative, interim, and summative assessments all provide data to be evaluated.

Online Delivery Systems

TLI. In the first comparison, TLI is an online vendor advocating to increase school districts' student performance. TLI provides curriculum alignment, interim assessments, research, consulting, technological services, and support to schools. TLI claims these services will aid teachers, administrators, and policy makers in meeting the needs of all students (A. Olsen, personal communication, April 6, 2010). School leaders in Arkansas are using TLI's services for assistance in assessment to help improve students' test scores on state assessments.

TLI may provide support for school districts regarding curriculum alignment and student assessment. TLI (2013a) stated that it focused on improving the academic performance of all students by applying a 3-level process that includes (a) examining, (b) aligning, (c) teaching and assessing curriculum, (d) reviewing assessment data, and (e) determining action based on the results. TLI provides assessments in literacy, mathematics, and science. These assessments provide the district information on strengths and weaknesses in both curriculum and student knowledge. Teachers and administrators may get immediate, daily feedback from the TLI web portal. TLI's curriculum staff consists of people with experience in the classroom. TLI provides curriculum support, professional development, and intervention strategies to elevate opportunities for all students to learn. The primary goal is that administrators, teachers, and students know the skills taught in the classroom as evidenced by continuous assessment. Curriculum alignment is vital to the program and curriculum maps are designed specifically to each district. The program reviews maps annually, and they may

be adjusted from year to year. These tools may provide helpful assistance for districts that use TLI.

TLI provides mathematics assistance. In the *TLI Mathematics Program*, TLI (2013f) said that it provides assessments in mathematics for Grades 1-8, Algebra I, Geometry, and Algebra II. Each assessment contains 20 multiple-choice questions and 1 open-response question. There are eight assessments for each level annually. The assessments follow the Common Core State Standards. TLI updates the tests annually.

TLI also provides assistance for literacy. In the *TLI Literacy Program*, TLI (2013e) noted that it provides assessments for Grades 1-11 in writing and Grades 2-11 in reading. The reading assessments contain 8 multiple-choice questions that focus on the reading passage and 1 open-response question. The writing assessment includes 12 or less multiple-choice questions based on grammar development. Assessments for Grades 1-5 have a writing prompt, and those for Grades 6-11 have the additional option to include a performance task. TLI supports two specific models for literacy. One design uses the Common Core State Standards for Grades 1-8 and uses the state frameworks and TLI's Core 14 skills for Grades 9-11. The second model focuses on Common Core State Standards for all grade levels.

Also, instructors and administrators will need training to enhance their skills so that students may receive the maximum benefit from the data provided by TLI. TLI (2013b) offers a broad range of professional development for educators. Training may range from web portal procedures to TLI reports targeted at improving education. Each new district to purchase the system receives professional development to train teachers for the interim assessment process. Once the primary teacher training has taken place,

appropriate school district employees may attend any of the advanced report trainings or curriculum based trainings. TLI provides Common Core State Standards training to help districts with the integration of the new set of standards. With Arkansas transitioning to the Partnership for Assessment of Readiness of College and Career assessment, integration of new standards will be paramount to educators promoting student achievement.

NWEA. In the second comparison, NWEA is used by many Arkansas school districts and has an origin from the western United States. NWEA (2013d) said that its origin dates back to 1974 when a partnership was formed between area schools in Portland, Oregon and Seattle, Washington. The coalition was created to improve the overall education system and create an opportunity for all students to learn. NWEA created one of the first computerized adaptive testing; and later, in 1977, established itself as a non-profit organization. In 1978, students in Portland and Hillsboro, Oregon participated in the first achievement level tests. The first computer adaptive assessments were administered to students from the Portland public schools in 1986. By 1997, the Measures of Academic Progress (MAP) was developed. MAP was a refinement of previous computer adaptive tests. In 2000, the first MAP tests were administered to 1,700 students. Beginning in 2001, the Learning Continuum was published by NWEA. NWEA created a partnership with more than 1,200 school districts in 2003. Within the next 2 years, NWEA introduced DesCartes: A Continuum of Learning and The Dynamic Reporting Suite. NWEA benefits from 5 decades of developing new tools and programs within the organization.

NWEA is a non-profit organization that offers many tools to analyze student learning. According to NWEA (2013a), it offers state aligned assessments, professional development for teachers, and numerous classroom resources which support proper evaluation of student learning. NWEA (2013e) stated that the its mission is “Partnering to help all kids learn” (p. 1). NWEA’s vision includes being the front runner in creating a child-centered education that uses comprehensive and accurate information to direct each student’s proper learning path. Information from the assessments is designed to inform and improve education. Professional development focuses on informing teachers of how to properly use the data and how the data directly guides the proper learning path for each student. NWEA administers a continuous effort to identify the most helpful data to improve student learning so that the information provided to teachers, parents, and administrators is most accurate and best aids each student. NWEA indicated that values are at the forefront of its program. NWEA’s core values include honesty, integrity, teamwork, continuous improvement, and care and respect for stakeholders. Further, they value accountability, rigorous data research, sustainability, and fun. NWEA (2013f) said that it has a belief in measurable growth and that all students may grow. Data-driven instruction and informed decision making lead the organization. NWEA contends that strong leadership and collaboration are essential for success. They credit the development of a learning culture and openness to new ideas and input as essential ingredients to their success.

Comparing the tools. Undoubtedly, there are many similarities between NWEA and TLI. Both of these assessments are performance level predictors for students taking the Benchmark and End of Course exams for the Arkansas Comprehensive Testing,

Assessment, and Accountability Program. Both of these testing companies are moving towards computer-based, multi-layered questions required by the Partnership for Assessment of Readiness of College and Career assessments. NWEA and TLI assessments provide student data that may be valuable to teachers in guiding instruction and focusing intervention.

In contrast, several fundamental differences exist between the two competitors. NWEA provides assessments classified as norm referenced assessments, which yield estimated positions in a predefined population, based on the measurement of a specific trait. NWEA provides a pre-test and a post-test. The test data from the NWEA assessment help identify gaps in training, and the DesCartes element shows the teachers what to teach to close the gap. The information is skill/mastery particular, due to the adaptive nature of the assessment. The reports generated following the assessments identify grade levels equivalency, academic growth, and Lexile ranges of independence for individual students. These reports may automatically group students based on specific emergent skills for targeted intervention.

Furthermore, TLI provides criterion referenced assessments, which use pre-determined cut-scores to conclude whether material has been learned by the students. These assessments are not adaptive and are specific to grade levels. Because of this, TLI build curricular units around these assessments. These formative assessments serve a similar function for student feedback as a chapter test would in a basal or other textbook.

Certainly, NWEA and TLI have many similarities and differences. However, both assessments provide districts, schools, departments, and classroom teachers with a myriad of data to target instruction and improve student learning.

Effects and Characteristics of Assessments

Proponents of formative assessment note that effective formative assessment may lead to effective summative assessment. Chappius and Chappius (2008) noted that the formative assessment is the act that delivers the information during the teaching process before the summative assessment. Guskey (2003) said that the best assessments to make a positive impact on student achievement are quizzes, writing assignments, tests and other assessments given daily to meet a certain set of standards. Wiliam (2006) expanded further by stating that educators must make deeper efforts and go beyond looking at just the assessments themselves. He added that school leaders should look at the real issues about how assessments support learners and their instructors. The data from the assessments should help them know where their students are, where they need to go, and how they may get there.

Two separate studies in the states of Virginia and Ohio identified the effective methods of schools within those states. In a study focusing on the Virginia schools, Demary, Magill, and Finley (2000) said that 22 of the 26 randomly selected schools listed assessment as an important effective practice. They said further that the most commonly mentioned activities were teachers making tests that aligned with the required state test and formatting the test. Another frequent practice was analyzing their classroom assessment data continually to assess and address the needs of the students. From the Ohio study, Kercheval and Newbill (2002) noted effective practices including the use of mirror tests and previous proficiency tests to familiarize students with the format. Simulation of the actual test day and the identification of students' strengths and

weaknesses revealed through data analysis were other noted practices. Formative assessment practices were evident in both studies.

Additionally, a research brief from the National Council of Teachers of Mathematics included an evaluation of the benefits of formative assessment. Wiliam (2007) said, “the effects of formative assessment observed in relatively short-scale studies are maintained over long periods of time” (p. 2). Wiliam said that classrooms where teachers were using short or medium cycle formative assessment processes produced approximately double the results in student learning compared to those classrooms not using formative assessment. Additionally, he identified greater engagement by students and greater professional satisfaction by teachers as positive effects of formative assessment. He noted that because effective formative assessment implementation is difficult, appropriate professional development for instructors was vital for the success of the process.

Also, a separate article was written examining the predictive validity of selected benchmark assessments used in the Mid-Atlantic region. Brown and Coughlin (2007) noted that, of the five commonly implemented test-preparatory assessments used in the region, only one provided evidence of predictive validity; and the data were limited. For the most part, success or failure on the online assessments did not translate to success or failure on the state exam.

Several studies uncovered characteristics of the different types of assessments. Perie et al. (2007) noted that desired characteristics of interim assessments for instructional purposes would include various question types along with multiple-choice. Provisions for qualitative insight about understanding and misconceptions, implications

for what to do besides re-teaching each missed items, high quality test items linked to intended standards, and a good fit with curriculum were essential to the process. They stated the analysis is an extension of learning and a good fit in pacing learning so students are not tested on untaught content. Vendlinski, Niemi, Wang, and Monempour (2008) noted, “The research literature and our experience suggests that scaffolding the assessment development process for teachers and providing a means whereby assessments can be continually polished should improve the quality of classroom formative assessments” (p. 1). Two ways to enhance the success and usefulness of formative assessment are identifying relevant test items and modifying pacing guides.

In October 2008, the Miami-Dade County Public Schools produced a report that analyzed the effect of interim assessment on student achievement in their schools. Blazer (2008) said, “Research is limited but indicates that when assessments are administered periodically throughout the school year and their results are used to guide instructions, they have a positive impact on student achievement” (p. 7). Further, she stated that research from studies indicates that the use of frequent testing with the added provision of corrected feedback leads to higher levels of student performance. These types of assessments appear to be particularly effective with lower achieving students. The program was implemented to improve the district’s instructional practices. The program’s three major components were (a) interim assessments, (b) a benchmark assessment item bank, and (c) professional development. Early indications suggested that the program has had a positive effect on the schools’ instruction and student achievement.

In recent years, the federal government has mandated summative examinations in mathematics and reading as part of No Child Left Behind (2002). A major reason for the

required testing is the need for improved academic instruction. From a study that identified a mastery learning intervention program in high school mathematics, Zimmerman and Dibenedetto (2008) said the program drew national recognition for its success. Teachers and students who were interviewed reported positive educational and motivational outcomes required of the mastery learning approach. Some characteristics of high stakes testing created concerns of drawbacks. A negative connotation was associated with these summative tests. However, Zimmerman and Dibenedetto said with proper formative assessment in the instruction model, the elimination of tests for the purpose of accountability would be possible. They said the model provides teachers with timely feedback about progress and need for improvement in meeting their goal as well as a curriculum that provides additional time and opportunities so all students may succeed. Shute (2008) stated identifying individual features of the learner as they relate to aspects of the tasks enhances the potential of using formative feedback. A teacher may get verification of response accuracy from the student, explanation of the correct answer, and use hints and examples to promote learning.

Another study that surveyed prospective secondary school teachers for England provided some assessment data. Taber et al. (2011) stated, “Assessment represents a complex area of educational thinking and practice, where professional learning may be challenging” (p. 182). He said that student interns may feel stress when analyzing discussion about what is good teaching versus their perception of assessment practices in secondary schools. He stated that although research supports that the formative assessment is most beneficial to student learning, pressures of high stakes testing alters instruction practices in schools.

Some studies support the relevance of effective interim assessments. From a paper presented at the National Council of Measurement in Education conference, Arter (2010) stated that some evidence exists to support interim assessments may result in increased student learning. However, the true value of any type of assessment comes from the information gained and used by both students and teachers. The assessments help them make useful formative and summative decisions. Further, she stated that it was not clear that redirecting resources used for interim assessment towards training teachers to be better classroom assessors would positively affect student learning. The benefit of the instructors' improved use of data were not clear. She said that all areas of assessment were important and that formative assessment, used properly, may create significant gains in student motivation and learning. Interim assessment results used primarily by teachers should be only a small part of the solution at best. Formative and summative assessments should be used to gather information in a balanced manner to inform all decision makers with data to meet all students' needs.

Effects and Characteristics of Curriculum Alignment

Many studies support the direct effect of curriculum alignment on student learning. According to a report from Appalachia Educational Laboratory (2005), research indicates a strong correlation between aligning the curriculum to the standards and assessment and student performance. Curriculum alignment involves textbook alignment, instructional alignment, alignment between the state standards and enacted curriculum, and alignment through professional development.

A study was conducted regarding effective practices in Virginia schools. DeMary, Magill, and Finley (2000) noted curriculum alignment was identified as an effective

practice 72% of the time. At least two interview groups in 18 of the 23 randomly selected schools identified curriculum alignment as important. They stated the three activities involving curriculum alignment revealed most frequently were focusing on the state required test, including all state examination content and an ongoing curriculum review.

The following study provided an even stronger correlation between curriculum alignment and student achievement. From the case study of effective methods for Ohio's improved districts, Kercheval and Newbill (2002) noted curriculum alignment was identified by the participants as the single greatest factor in achieving improved student test scores. The Delphi portion of the study ranked curriculum alignment as the most important component and the most frequently reported factor. They added from the available data that two themes consistently emerged relating to curriculum alignment: curriculum mapping and change in instructional practices. Mapping and instructional changes may empower teachers to improve the effectiveness of their teaching practices.

Research on curriculum alignment provides support that the alignment may improve student performance. Squires (2012) suggested the student achievement be related to three alignment categories: (a) the taught curriculum, (b) the tested curriculum, and (c) the written curriculum. Pairing any of these categories produces a positive impact for alignment. Alignment of the state standards to the curriculum is vital for a student's understanding of the content. A teacher's instruction must include tested material or students will not understand the necessary content. He noted that standards are considered to be a more general list of learning expectations, and curriculum is considered to be more specific. He said that student writing assignments should each be aligned to a specific standard. Polikoff (2012) said "recent research suggests that

instructional alignment tends to be higher when the standards and assessments are more aligned” (p. 364). Further, while legislators consider strengthening standardly based reforms through Common Core State Standards, it is essential that close attention be paid to align resources of instructional influence.

Many factors are deemed as important in their relationship to student achievement. However, Johnson (2013) indicated that an efficiently aligned curriculum could off-set other traditional indicators of student achievement like gender, socioeconomic status, ethnic origin, and teacher affect. She advocates three processes that are necessary for improving curriculum alignment: basic alignment services, knowledge domain alignment, and comprehensive alignment. Any or all of these methods may be needed depending on the needs of the school district’s curriculum.

Research identified vertical alignment as an important component for the promotion of student achievement. A comparison was conducted between the vertical alignment of the curriculum in education with the wheel alignment for a vehicle (Holmes, 2012). It said that the wheel alignment minimizes the amount of tire wear and keeps the car from pulling to one side. The mechanic must adjust the angles of the wheels to maximize performance of the vehicle’s wheel alignment. Likewise, vertical alignment of the curriculum ensures that the teacher’s instruction is directly aligned to state standards and assessments. Vertical alignment enhances the likelihood that what is being taught and tested in the classroom matches directly with the state standards and assessment. Vertical alignment should be a district-wide process that requires collaboration from teachers, curriculum specialists, and administrators to ensure a meaningful, accurate curriculum is in place in the learning environment.

Vertical alignment is a component that is commonly referenced in relation to improved student performance. Two districts attributed their implementation plan for vertical alignment to the upgrade in their students' academic performance on Benchmark Mathematic and Literacy Exams. The districts' specifically cited the teamwork of district and school officials and utilization of the curriculum mapping software as the major components of their vertical alignment success. Kurz, Talapatra, and Roach (2012) said, "to ensure coherency among all curricula of an accountability system, horizontal alignment between intended and assessed curricula must be extended vertically to curricula at the teacher and student levels" (p. 43). Alignment practices must be executed properly to support teacher instruction that will lead to optimum student performance.

Hypotheses

The following hypotheses were generated for this study:

1. No significant differences will exist by gender in Benchmark Literacy Exam performance of sixth-grade students in Arkansas middle schools who received TLI-interim assessment versus those sixth-grade students in Arkansas middle schools who received NWEA-interim assessment.
2. No significant differences will exist by gender in Benchmark Literacy Exam performance of those seventh-grade students in Arkansas middle schools who received TLI-interim assessment versus those seventh- grade students in Arkansas middle schools who received NWEA-interim assessment.
3. No significant differences will exist by gender in Benchmark Mathematics Exam performance of sixth-grade students in Arkansas middle schools who

received TLI-interim assessment versus those sixth-grade students in Arkansas middle schools who received NWEA-interim assessment.

4. No significant differences will exist by gender in Benchmark Mathematics Exam performance of seventh-grade students in Arkansas middle schools who received TLI-interim assessment versus those seventh-grade students who received NWEA-interim assessment.

Description of Terms

Benchmark assessments. Coffey (2010) defined benchmark assessments as short examinations given throughout the school year that give teachers immediate feedback on how their students are meeting state standards. These assessments are considered a tool to measure student growth and to adjust curriculum to meet specific learning needs.

Common Core State Standards. In their *Mission statement*, the Common Core State Standards Initiative (2012) defined the Common Core State Standards as the current curriculum adopted by 45 states designed to provide a clear, consistent expectation of the material that students are expected to learn. They provide the blueprint of what teachers and parents are to do to help them. The standards are designed to be relevant to the real world and arm students with the knowledge and skills that they need to be successful in college and careers.

Curriculum alignment. According to Appalachia Educational Laboratory (2005), *curriculum alignment* is defined as the extent to which textbooks and instruction are aligned with standards and assessments.

Formative assessment. Brown and Coughlin (2007) defined formative assessment as assessment designed to provide data to guide instruction.

Formative feedback. Shute (2008) defined *formative feedback* as information communicated to the student with the intention to modify his or her thinking in order to improve the student's learning. It is usually presented to the student in a non-evaluative manner as information in response to the student's work.

Horizontal alignment. In their *Horizontal Alignment Definition*, EHow (2013) defined *horizontal alignment* by stating, "In education, horizontal alignment is defined as the alignment of the curriculum being taught by teachers in a common grade level. It provides teachers with a guide for their instruction" (p. 1). Horizontal alignment is sometimes referred to as a pacing guide.

Interim assessment. Perie et al. (2007) defined *interim assessment* as assessment designed to evaluate students' knowledge and achievement within a specific time frame to a specific set of academic goals and designed to inform decisions in the classroom and beyond.

No Child Left Behind. The No Child Left Behind of 2001 (2002) federal legislation was an enactment of a set of theories for standards-based reform in education. It centered on the belief that establishing higher standards based on measurable goals may improve education in America. It focused on common expectations for all students which would reduce the racial, class gaps.

Professional development. Reeves (2006) defined *professional development* as research-based training designed to help improve student success. The training should be focused on classroom feedback, assessment practices, and cross-disciplinary writing for it to reach maximum effectiveness.

Summative assessment. Perie et al. (2007) defined *summative assessments* as assessments given one time at the end of the semester or school year to evaluate student performance against a defined set of standards.

Vertical alignment. *Vertical alignment* is defined as the process that creates a seamless flow of instruction from grade level to grade level (“Vertical alignment,” 2013). It was designed to minimize wasted time on excess review and focus on the material deemed most necessary.

Significance

Research Gaps

Research studies determining the effect of new instructional strategies on student performance have provided valuable data to educators throughout the course of educational history. Background information for this study uses research data to connect assessment and alignment to student achievement. However, new research that may connect the significance of online programs towards student success may provide valuable information, which was not previously available. Specifically, data providing insight on the effectiveness of TLI and NWEA programs will be valuable to K-12 educators. The knowledge could impact decisions for school districts that may directly affect student performance.

Potential Implications for Practice

Determining the effect of TLI interim assessment on Benchmark scores compared to Benchmark scores from NWEA-interim assessment may provide valuable data that may impact school leaders’ decisions, which have the potential to affect student learning. For instance, whether or not the TLI program or the NWEA program has an effect on

Benchmark scores will help determine whether or not the investment is benefiting teacher instruction and student success.

If TLI-interim assessment program or NWEA-interim assessment program instruction has a more positive effect on Benchmark scores, the school will benefit from gaining this information. The data accumulated might help them assess if their investment in the program is improving their standing for state accreditation in meeting the standards. If neither the TLI program nor the NWEA program is effective, schools may look at other resources to assist their efforts to improve student scores on Benchmark exams.

Additionally, another major implication to consider is the adaptation of TLI and NWEA to the transition from the traditional state standards and Benchmark testing to the Common Core State Standards and Partnership for Assessment of Readiness of College and Career testing. The recent shift in curriculum and testing in Arkansas and other states should be important to monitor and prepare for essential adjustments.

Process to Accomplish

Design

A quantitative, causal-comparative strategy was employed in this study. The independent variables for all four statements of the problem were interim assessment (TLI versus NWEA) and gender. The dependent variable for Hypothesis 1 was the achievement measured by the Benchmark Literacy Exam for sixth-grade students. The dependent variable for Hypothesis 2 was the achievement measured by the Benchmark Literacy Exam for seventh-grade students. The dependent variable for Hypothesis 3 was the achievement measured by the Benchmark Mathematics Exam for sixth-grade

students. The dependent variable for Hypothesis 4 was the achievement measured by the Benchmark Mathematics Exam for seventh-grade students.

Sample

The study used sixth and seventh-grade students in four school districts' middle schools in Northwest Arkansas. Two districts' schools used the TLI program, and two districts' schools used the NWEA program. Students were categorized based on which test they were taking. Sixth-grade students were used for the sixth-grade Benchmark Literacy and Benchmark Mathematics Exams. Seventh-grade students were used for the seventh-grade Benchmark Literacy and Mathematics Exams.

The four school districts were chosen based on their similarity in size of the school, student demographics, and ethnicity. The first TLI district had 306 sixth-grade students and 321 seventh-grade students. Of the sixth-grade students, 51% were females, and 63% of the students were Caucasian. Of the seventh-grade students, 50% were females, and 67% of the students were Caucasian. The second TLI district had 1,135 sixth-grade students and 1,010 seventh-grade students. Of the sixth-grade students, 50% were females, and 54% of the students were Caucasian. Of the seventh-grade students, 50% were females, and 56% of the students were Caucasian.

The first NWEA school district has 680 sixth-grade students and 691 seventh-grade students. Of the sixth-grade students, 47% were females, and 69% of the students were Caucasian. Of the seventh-grade students, 49% were females, and 70% of the students were Caucasian. The second NWEA school district had 1,509 sixth-grade students and 1,501 seventh-grade students. Of the sixth-grade students, 48% were females, and 40% of the students were Caucasian. Of the seventh-grade students, 48%

were females, and 39% of the students were Caucasian. All of the selected school districts were within a 50 mile radius of one another.

The first two school districts used the TLI-interim assessment program. The second two school districts used the NWEA-interim assessment program. Teachers from the first two school districts were trained to use TLI software with appropriate professional development. Teachers from the second two school districts were trained to use the NWEA software with appropriate professional development. These teachers used the preparatory assessments provided by TLI or NWEA and aligned their curriculum for the courses with the software. The training occurred for the teachers prior to their instruction for the school year. Samples were taken from the following populations: males receiving TLI-interim assessment (first two schools), males receiving NWEA-interim assessment (second two schools), females receiving TLI-interim assessment (first two schools), and females receiving NWEA-interim assessment (second two schools).

Instrumentation

In the fall semester, students at the first two school districts were instructed in sixth- and seventh-grade literacy and mathematics using TLI-interim assessment. Students at the second two school districts received NWEA-interim assessment program in sixth- and seventh-grade literacy and mathematics. Sixth- and seventh grade students at the four school districts took the Benchmark Literacy and Mathematics Exams in the spring term.

Data Analysis

To address the first hypothesis, a 2 x 2 factorial was conducted using the condition by gender as the independent variable and the raw scores for the sixth-grade

Benchmark Literacy Exam the dependent variable. To address the second hypothesis, a 2 x 2 factorial was conducted using the condition by gender as the independent variable and the raw scores for the Seventh-grade Benchmark Literacy Exam the dependent variable. To address the third hypothesis, a 2 x 2 factorial was conducted using the condition by gender as the independent variable and the raw scores for the Sixth-grade Benchmark Mathematics Exam the dependent variable. To address the fourth hypothesis, a 2 x 2 factorial was conducted using the condition by gender as the independent variable and the raw scores for the Seventh-grade Benchmark Mathematics Exam the dependent variable. To test the null hypothesis, a two-tailed test with a .05 level of significance was used.

CHAPTER II

REVIEW OF THE RELATED LITERATURE

Today's school leaders are actively looking for ways to accelerate student academic growth. Online delivery systems are one avenue that educators are exploring to address improving student achievement. With concerns of reduced funding for education and pressures of academic accountability on student test scores, educators must make informed decisions before purchasing services from these online providers. Axelson (2005) stated that desire for online formative assessment providers is based on the need for educators to receive accurate, meaningful data both quickly and regularly for planning instruction and intervention.

TLI and NWEA were the online delivery systems representing the focus of this project. Both providers offer resources to help provide both formative and interim assessments to help teachers prepare students for their Benchmark Exams. The state Benchmark Exams were the summative assessments that measured student achievement. Additionally, TLI and NWEA provided data to help guide curriculum alignment ensuring that students were tested over what they were taught. TLI and NWEA are currently used by many Arkansas schools to help districts address the academic needs of students.

Besides interim assessment programs, gender was another important component to consider. The effect of gender on student achievement has been researched extensively through time. Gender perceptions have existed for centuries regarding academic

capability. Are they myths, or do they have some validity? In this study the text data by interim assessment providers and gender were compared. The information gained from this study may guide the decisions of school administrators and educators.

Online Interim Assessment Programs

School districts from Arkansas and other states are turning to online providers for assessment and curriculum alignment. The need for such services has been influenced by pressures on school districts to improve student test scores that affect the school's accountability measures. In Arkansas, TLI and NWEA are the two main online interim assessment programs that school leaders use. This study focused on their effectiveness in improving student test scores.

School leaders have considered purchasing interim assessment programs to help their students prepare for state summative assessments. The federal legislation of No Child Left Behind (2002) increased the pressure on public schools to push their students to perform well on the state exams. No Child Left Behind was legislation that enacted a standard-based reform in education centering on establishing high standards based on measurable goals and common expectations for all students. The intent of the legislation was to improve education in America and reduce the racial and class gaps.

In reaction to No Child Left Behind, individual states adjusted their accountability systems to meet the new standards. In Arkansas, each school had to meet a sliding percentage scale of proficient students in literacy and mathematics for certain grade levels in order to be considered an adequate school; the percentage was ultimately raised to 100%. To measure student proficiency, Arkansas used Benchmark Exams starting in elementary school through Grade 8. For students' high school years, the schools used the

End of Course Literacy (Grade 11), Algebra, and Geometry examinations to measure student performance. Schools failing to reach the required percentage of proficient students would be placed on alert status. If they continued to not attain the goal, they would be placed on school improvement status. Eventually, they would be placed under the state takeover if they did not reach the goal.

A negative academic designation of a district's school has several potentially negative effects. A district might lose students to a higher performing school because parents want their children to have a better chance to succeed. Furthermore, a district might struggle to find competent teachers because of the stigma connected with poor academics. Ultimately, state and federal funding might be affected because of a failure to produce the required results, not to mention state takeover of the school. Thus, school leaders are motivated to improve test scores, and an interim assessment program becomes an appealing tool to help them meet state and federal academic requirements.

When determining whether to purchase an interim assessment program, an educator should consider the expectations of the program and if an interim assessment program will benefit the students' academic achievement and growth. Generally, programs provide assessment data measuring different levels of achievement. Fleming (2013) stated that it is essential to have an understanding of students' current instructional level to tailor instruction to each student's academic needs and strengths. Thus, one potential benefit of an interim assessment program would be improved instruction.

An interim assessment program would be most beneficial if it could be tailored specifically to each district's student population. For instance, Ryan (2010) stated, "The complexities and demands for validity across a multi-component assessment system

designed for a diverse student population are critical and considerable” (p. 17). Attention must focus on validity for all students with finances, materials, and personnel. Efforts focusing on reviewing current practices and discovering valid approaches to developing and maintaining an effective assessment system must also be in place. A consistent, systematic evaluation system of the program will be essential to justify the investment in the assessment system. Further, the interim assessment system that matches the needs of the students of the district will provide the most benefit.

Another interesting point of view was found from a study exploring the use of interim assessments in classroom instruction. Goertz, Olah, and Riggan (2009) stated that, although their study showed that interim assessments were useful, they were not sufficient standing alone to inform instructional improvement. However, the information from the interim assessments did aid teachers in decisions over what content should be re-taught and to which students. The data helped identify specific areas that the class as a whole performed well on, as well as specific areas in which students did poorly. Both school and district factors influenced the use of interim assessments for aligning interim assessment content with standards and curriculum and for developing an expectation that the results would be used to inform classroom instruction. The factors also helped provide instructional support for students needing additional assistance and support to instructors for data analysis and instruction. They stated, “We found little evidence, however, that the interim assessments we studied helped teachers develop a deeper understanding of students’ mathematical learning—a precursor to instructional improvement” (p. 238). Summarizing, the study found many positive aspects of the use

of interim assessments. However, other factors like school leadership and culture of data were also critical in addressing the improvement of instruction.

Improving instruction and determining student needs are important potential benefits of purchasing an interim assessment program. However, are there other benefits to consider? Hicks (2014) devised several points to consider when comparing and evaluating interim assessments. First, the purchaser should consider whether the assessment is designed to provide achievement status and growth data for students and if the assessment is adaptive. Second, he stated that it was imperative the assessment link appropriate resources and inform decision-making for classroom instruction. In other words, can the data be used to influence decisions on the building and district levels? Third, the data should also be valid and be able to predict outcomes of high-stakes summative tests, as well as college entrance exams. Several potential benefits should exist before purchasing an interim assessment program. School leaders must be clear about the program expectations to make the best decision on the purchase.

The cost of a program is an extremely important consideration when choosing to purchase an interim assessment system. The access and availability of finances vary from district to district. Further, administrators have to justify the gain in student academic growth versus the sometimes-significant financial investment. Ryan (2010) explained, “A systematic evaluation of the overall impact of the assessment system is needed to account for the massive investment in the enterprise” (p. 19). Therefore, a value analysis is essential prior to the commitment to an interim assessment program.

TLI and NWEA are currently used in the state of Arkansas. Local curriculum specialists were questioned to get an idea about the cost of each program. The cost of TLI

is approximately \$25 per student. An additional cost of time was noted for those that used TLI with teachers printing, organizing, and grading writing and open-response questions. NWEA ranges from \$12.50 to \$13.50 per student, with an option for science for an additional \$2.50 per student (S. Owens, personal communication, December 1, 2014). Consequently, the estimated cost of TLI is approximately twice as much as that of NWEA. If both programs are equal in effectiveness in promoting student success in literacy and mathematics, NWEA's price would seem to give it the advantage.

TLI is an Arkansas-based company that has provided interim assessments for helping educators improve instruction by providing valuable data promptly. TLI (2013c) stated, "TLI provides a single point of access to all district achievement data so that teachers and district leaders can have the right information—right now—to make the right decisions for their students" (p. 1). Further, student tests, answer sheets, and reports may be accessed through the web portal. Additionally, the web portal provides document libraries with lesson plans, other presentation materials, and reports from mathematics, literacy, and science interim assessments in conjunction with the districts' Benchmark state testing data.

TLI provides interim assessments for both literacy and mathematics. TLI (2013e) stated that their literacy interim assessments provides for writing and reading. Writing assessments are provided for Grades 1-11, and reading assessments are provided for Grades 2-11. Further, the reading assessments include eight multiple-choice questions based on the reading passage and one open-response. Writing assessments for Grades 1-5 include 12 multiple-choice questions and 1 writing prompt. Writing assessments for Grades 6-11 have 12 multiple-choice questions and use a performance task. Four reading

interim assessments and four writing interim assessments are provided in an academic year. TLI curriculum specialists must approve all reading passages, reading items, and writing items before assessments are used. Likewise, TLI (2013f) stated that interim assessments are developed for Grades 1-8 in mathematics and for Grades 9-12 in Algebra I, Geometry, and Algebra II. Each interim assessment provides 20 multiple-choice questions and 1 open-response question. There are eight assessments provided per yearlong course, and TLI specialists approve the content and standard correlation.

For literacy and mathematics, TLI assesses students against the Common Core State Standards aligned by the district (Common Core State Standards Initiative, 2012). The Partnership for Assessment of Readiness of College and Career testing, aligned to the Common Core State Standards, began with the spring of 2015. Consequently, TLI (2013d) noted, “New standards require a reevaluation and revision of current approaches to both instruction and assessment” (p. 1). Further, TLI claimed to provide improved tools to make the transition easier for students and teachers. TLI’s full transition plan is valuable information for curriculum directors and other school leaders going forward.

NWEA was created in the 1970s by an alliance of Washington and Oregon educators who wanted to create a new type of testing system. In *Our History* (NWEA, 2013d), the authors stated that the creators of NWEA wanted “to create a new, precise way to measure an individual student’s academic level and growth” (p. 1). Further, NWEA wanted the resulting data to serve as a transforming teaching tool. NWEA is currently being used in many states, including Arkansas, for interim assessment.

NWEA offers interim assessments for many grade levels. The MAP tests for Primary Grades are being used by districts around the United States. The MAP is used for

grades 3-12, and the MAP tests for Primary Grades are used for Grades K-2. The NWEA (2013b) authors noted that the MAP generates a personalized learning experience that adapts to each student's learning level. Assessment data are available within 24 hours of the testing, revealing what the student knows about different areas of study, as well as what the student is ready to learn in those areas. The MAP includes assessment in reading, language arts, and mathematics. The tests are untimed, but an average student typically completes the assessment in less than an hour per subject. These assessments are given 3-4 times per year.

With Partnership for Assessment of Readiness of College and Career testing on the horizon for Arkansas public schools, the transition to Common Core State Standards is at the forefront of educators' minds when considering an interim assessment program (Common Core State Standards Initiative, 2012). NWEA is attempting to make strides in that transition. Hicks (2014) suggested four ways that MAP or the MAP tests for Primary Grades interim assessments could support the transition to Common Core. First, MAP measures both student growth and instructional readiness at the start, middle, and end of the year. Second, MAP assessments provide growth information to provide context for data interpretation and evaluation for after the transition. Third, MAP offers a data point for measuring college and career readiness for all students. Fourth, MAP allows teachers to evaluate the program's effectiveness. For Arkansas educators, an interim assessment's ability to adapt from the past curriculum and assessment state requirements to the rigor of Common Core State Standards and Partnership for Assessment of Readiness of College and Career testing could be one of the deciding factors in purchasing a program.

Assessment and Curriculum Alignment

Assessment and curriculum alignment serve as the major benefit of using an interim assessment system. Curriculum alignment organizes the material that is to be taught. Appalachia Educational Laboratory (2005) defined curriculum alignment as the extent to which instruction and textbooks are aligned with standards and assessment. As mentioned, interim assessment and curriculum alignment are two tools provided by interim assessment providers. The effects of assessment and alignment on student test scores are an important consideration in the present study.

When considering the value of an interim assessment program, curriculum alignment is a vital factor to evaluate. The academic effects of organizing curriculum and instruction have been a popular topic of educational research. Squires (2012) indicated that student achievement is directly related to taught curriculum, tested curriculum, and written curriculum. Moreover, the pairing of any of the three may have a positive effect on alignment and, ultimately, student achievement. Additionally, Johnson (2013) stated that effective curriculum alignment might potentially offset other traditional indicators of student achievement such as gender, socioeconomic status, ethnic origin, and teacher effect.

Curriculum alignment's value is referenced in many areas concerning student achievement. Brown and Green (2014) identified curriculum alignment as one of the seven leadership strategies in the literature of school transformation. Specifically, they stated that "...the leadership practice of monitoring students' learning and using the results to plan individualized instruction was the highest determining factor of all the

practices that led to student transformation” (p. 13). This involves both assessment and alignment.

Curriculum alignment’s educational benefits may affect student achievement both directly and indirectly. Alignment may have a positive effect on instruction. For instance, Polikoff (2012) stated that instructional alignment was more effective when standards and assessments were well aligned. Thus, proper curriculum alignment affects instructional alignment, which might improve instruction. The value of alignment should not be overlooked. As mentioned above, it has a potentially profound direct and indirect effect on student achievement.

The effect of interim assessment on summative assessments is an important factor to consider as school leaders today attempt to improve student achievement in their districts. If interim assessments affect the standardized scores positively, then administrators have to consider their value to the administrators’ districts. A study was conducted examining the effects of a commercial interim assessment program used in Indiana public schools. The program’s effects on reading and mathematics were examined. Konstantopoulos, Miller, and Ploeg (2013) stated that the results indicated that the treatment effects were positive; however, the results were not consistently significant. Furthermore, the treatment effects were greater in Grades 3-8 and smaller in Grades K-2. The results in mathematics for Grades 3-6 showed the most significant treatment effects. Additionally, McGraw Hill Education (2015) reviewed this study and endorsed the findings. They stated that the study revealed significant positive effects in both mathematics and reading. Specifically, mathematics scores were most significant in Grades 5 and 6, and reading gains were most evident in Grades 3-5. Summarizing, the

results of the Indiana study indicated that the use of an interim assessment program did have a positive effect on student literacy and mathematics achievement.

Conversely, a study reviewing the influence of commercially-prepared interim assessments on math and language arts achievement for an eighth-grade student in New Jersey showed different results. Babo, Tienken, and Gencarelli (2014) stated that the study revealed a lack of quantitative data to determine the true value of these interim assessments on school reform. Examining a commercially-produced assessment program, the study explored the program's effects on mathematics and literacy using pre- and post-examinations. The researchers also examined the program's ability to predict results on state-mandated exams on those subjects. They stated that there was no significant difference in the pre-test predictive ability to that of the post-test predictive ability. Therefore, the efficacy of the implementation of the interim pre-tests and post-tests was brought into question. Blazer (2008) also indicated uncertainty on the direct relationship between interim assessments and student achievement. She stated, "Although some interim assessments have predicted students' subsequent performance, there has been much variability in the magnitude of these relationships" (p. 8). Therefore, the studies explored were mixed concerning the value of interim assessment programs.

Both TLI and NWEA promote curriculum alignment. TLI (2013b) stated that curriculum alignment is valued as the first and possibly the most important step in the continuous improvement model. They indicated that alignment to each subject is slightly different. However, the alignment should provide a roadmap that connects learning expectation in the classroom to the test items on the interim assessments. Similarly, NWEA (2013c) stated that their standards' alignment process guarantees that each item

reflects standard content precisely. They ensure that the flexibility to maintain current standards is adjusted to new standards based on the district's state requirements. Thus, both interim assessment providers have areas on their website dedicated to curriculum alignment.

Gender

When considering interim assessment, gender biases that may exist must be taken into account when evaluating the assessments. Throughout the years, many theories have been discussed regarding the effect of gender on academics. A perception exists that boys have a better chance to be successful in mathematics compared to girls. Further, another perception exists that girls have a better chance for success in literacy concepts, reading, and language arts when compared to boys. This section examines what the research says regarding gender perceptions; particularly, actual effects of gender on literacy and mathematics achievement will be reviewed.

Stereotypes associating academic success or lack of academic success have existed for a long time. These perceptions have an effect on how much time and what type of effort children devote to certain academic activities. Therefore, these overt perceptions could positively affect females in literacy efforts and males in mathematics efforts. Thus, the stereotyped perceptions could affect females negatively in mathematics and affect males negatively in language arts. Research indicates that parental perceptions of these stereotypes might have an effect on the parents' expectations of their children's success. Eccles, Jacobs, and Harold (1990) analyzed the effect of parents' expectations in influencing their children to participate in gender role stereotyped activities. They surmised that parents "distort their perceptions of their children in gender-role stereotypic

activities such as math and sports” (p. 183). They found that parents could have a strong positive or negative effect on their students’ performance in gender-role stereotypic activities. These perceptual biases influence students’ self-perceptions and choices of activities in which to participate. They stated, “parents’ perceptions of their children’s competencies in math, English, and sports are strongly related to independent indicators of their children’s actual competence in these domains” (p. 197). The parents’ perceptions were influenced by stereotypical gender beliefs. The evidence supported that influence is independent of any actual difference in children’s competencies. They believed the parents’ perception had an effect on the interest and confidence in participating in these activities. In summary, the perception of the parents, not the student’s actual ability, was determined to have an effect upon the students.

Gender role stereotyping was investigated through a study that focused on Kenyan primary schools. Ogutu, Peter, and Collins (2011) stated that the results from their study indicated a feminine and masculine perception existed in certain subjects. Male students had aspirations for mathematics and science, and female students favored language arts. Furthermore, boys’ preference towards mathematics was attributed to more opportunities to manipulate numerical objects that exposed them to problem-solving skills. Girls had less exposure to activities that enhance problem-solving skills, but they were given more strenuous responsibilities at home. Girls’ preference to language arts was attributed to their interest in social context and communicative tasks, and boys were more interested in technical matters. Stereotyping has an effect on male and female interest in subject areas. Actual ability and potential being more suited to gender in the subject areas was not found.

A study was conducted on a sample of 342 senior secondary students. A survey was provided that included 20 stereotype self-concept questions and 10 mathematics questions. Igbo, Onu, and Obiyo (2015) stated, “The findings of the study indicate that gender stereotype has significant influence on students’ self-concept and academic achievement in favor of male students” (p. 2). Further, they stated that school location influenced academic achievement but not students’ self-concept. For the study’s purposes, stereotype means “segregation between one group of people and the other showing superiority over one group against the other” (p. 16). From the study’s findings, gender stereotype blocked female students from the rights and privileges that they were entitled to because they were intentionally segregated from certain activities that were viewed as male-oriented activities. Summarizing, the results showed that gender stereotype had significant influence on academic achievement, but self-concept and location had no significant influence on the students in the study. Again, gender stereotyping is found to have an effect on student learning, but not gender itself.

Research on gender’s effects on literacy achievement has derived another conclusion. Zembar and Blume (2011) attributed part of the explanation of gender differences in literacy achievement to the cognitive abilities of middle school students. Further, female students outscore male students in verbal reasoning, verbal fluency, comprehension, and understanding logical relations. They stated that girls, by the age of 12, began to like language arts and social studies more compared to boys. Because interest is closely linked to confidence, a student’s confidence in a subject may have an effect on how well that student does in a subject area.

Research provides many examples of girls out-performing boys in literacy achievement. Klinger, Shuhla, and Wade-Woolley (2009) stated that female students had been shown to have a significant and consistent advantage in literacy achievement over their male counterparts in North America and across the world. However, the differences vary and, in some cases, literacy achievement between the genders is similar. In schools where the gap was less, boys reported higher levels of fiction reading, reading frequency, and homework completion. Boys played video games less frequently in the low-gap schools. Good teaching and positive learning environments were listed as constant factors in the low-gap schools.

As with literacy, research studying the effects of gender on mathematics achievement has yielded many results. Zembar and Blume (2011) stated that, generally, males score higher compared to females on mathematics achievement exams. However, males and females perform equally well on areas of basic mathematics knowledge. Girls even perform better compared to boys in mathematics computational skills. Conversely, gender differences become relevant to the areas of mathematical reasoning and geometry where boys greatly outperform girls. The researchers noted that males exhibited a much higher confidence level in their mathematics skills compared to their female counterparts. Studies have consistently shown that factors such as self-confidence and stereotyping are listed more than any inherent skills of males or females when it comes to academic achievement.

Many times, gender differences in mathematics achievement are believed to emerge in late middle school and early high school. However, Penner and Paret (2008) stated that gender differences in mathematics achievement begin when children are much

younger. In fact, they observed differences as early as kindergarten. The male advantage at the earliest ages was attributed to parents' high level of education. In addition, all males tended to do as well or better than girls do by the third grade in mathematics.

Some reports indicate that gaps in gender achievement are closing. Hyde and Mertz (2009) indicated that girls are now achieving at an equal level with boys in American high schools. Further, female students are excelling in high schools that had significant gender gaps decades before. Though some research shows that boys outperform girls in mathematics in school, factors of perception, confidence, and parental influence seem to influence that difference and not inherent factors of males and females. Consequently, some research indicates that gender does not have a significant effect on academic performance at all. Ebeonuwa-Okoh (2010) conducted a study that examined the influence of age, financial status, and gender on academic success. The research indicated that none of these factors, including gender, were significant indicators for academic performance. They found that students' character and behavior were the foremost determinants of academic performance. Although some studies indicate that gender and gender perceptions affect students in some academic areas, the reviews are clearly mixed.

Conclusion

School administrators have pressure for their schools to achieve a certain standard based on state accountability measures. Interim assessment programs might be able to provide assistance in improving the academics in public schools. School leaders must weigh the cost of the investment versus the potential gains in making a decision whether to purchase the product. From the research, interim assessment has the potential to have a positive effect on student achievement. Although the research showed mixed results on

the success of online interim assessment programs in influencing student achievement, the current research study should provide more information on the educational topic of interim assessment programs. TLI and NWEA both provide curriculum alignment guidelines, and research shows that the organization of curriculum and instruction may affect student achievement.

Perceptions that males excel in mathematics and girls excel at literacy have existed for some time. Results of research showed mixed results on the topic. In addition to looking at Benchmark Exam results comparing interim assessment providers, the juxtaposition by gender was also explored in the current study. The test results of sixth and seventh-grade boys and girls were reviewed providing data for gender comparison in the subject areas of mathematics and literacy.

CHAPTER III

METHODOLOGY

Some research on interim assessment systems indicates that proper use of data from these assessments may benefit students and teachers. However, effective use of assessment data is based on several significant assumptions. School districts must communicate strong expectations for the data from interim assessments to guide instruction. The assessments must generate valid information on student understanding of the tested content. School leaders and instructors must have the knowledge to interpret assessment results. Technology will streamline the entire process (Goertz et al., 2009). Further, Ryan (2010) stated, “Predicting students’ likely performance on important future activities is a valued application of interim assessment data and, with caution and care, procedures can be explored to develop and apply appropriate prediction models” (p.10). As is evident by the review of literature, more research is needed in the area of interim assessment systems and their potential effect on student achievement. Interim assessment providers are becoming more plentiful and attempting to determine the most effective programs may be valuable educational research. This study will address the effectiveness of two frequently used interim assessment providers being used in Arkansas and attempt to compare their effectiveness on state Benchmark Exams in both literacy and mathematics.

The study examined the effects of each interim assessment provider and gender on Benchmark performance in literacy for sixth-grade students. Second, the study examined the effects of each interim assessment provider and gender on Benchmark performance in literacy for seventh-grade students. Third, the study examined the effects of each interim assessment provider and gender on Benchmark performance in mathematics for sixth-grade students. Fourth, the study examined the effects of each interim assessment provider and gender on Benchmark performance in mathematics for seventh-grade students.

The research hypotheses are as follows:

1. No significant differences will exist by gender in Benchmark Literacy Exam performance of sixth-grade students in Arkansas middle schools who received TLI-interim assessment versus those sixth-grade students in Arkansas middle schools who received NWEA-interim assessment.
2. No significant differences will exist by gender in Benchmark Literacy Exam performance of those seventh-grade students in Arkansas middle schools who received TLI-interim assessment versus those seventh- grade students in Arkansas middle schools who received NWEA-interim assessment.
3. No significant differences will exist by gender in Benchmark Mathematics Exam performance of sixth-grade students in Arkansas middle schools who received TLI-interim assessment versus those sixth-grade students in Arkansas middle schools who received NWEA-interim assessment.
4. No significant differences will exist by gender in Benchmark Mathematics Exam performance of seventh-grade students in Arkansas middle schools who

received TLI-interim assessment versus those seventh-grade students who received NWEA-interim assessment.

Research Design

A quantitative, causal-comparative strategy was selected for the study. The independent variables for Hypothesis 1 were type of interim assessment and gender. The dependent variable was the Grade 6 literacy achievement measured by the Arkansas Benchmark Literacy raw score. The independent variables for Hypothesis 2 were type of interim assessment and gender. The dependent variable was the Grade 7 literacy achievement measured by the Arkansas Benchmark Literacy raw score. The independent variables for Hypothesis 3 were type of interim assessment and gender. The dependent variable was the Grade 6 mathematics achievement measured by the Arkansas Benchmark Mathematics raw score. The independent variables for Hypothesis 4 were type of interim assessment and gender. The dependent variable was the Grade 7 mathematics achievement measured by the Arkansas Benchmark Mathematics raw score. The design of the study was used to isolate groups who received the same interim assessment systems for the school year. With the quantitative, causal-comparative strategy, an ex-post facto design was practical because the event had already occurred, therefore, the data were already available. Each of the four hypotheses employed a 2 x 2 factorial between-groups design. A weakness of the causal-comparative strategy is that these studies yield weaker results determining evidence for causation than experimental studies.

Sample

The sample for this study consisted of 7,153 sixth- and seventh-grade students. The students were from Northwest Arkansas middle schools and took the mathematics and literacy Benchmark Exams for their grade. Four school districts were selected. Two school districts used TLI-interim assessment program. Two school districts used the NWEA-interim assessment programs. A comparison sample was drawn from each of these school districts. All of the schools had similar demographic characteristics. All students for the study were students taking the sixth-grade mathematics and literacy Benchmark Exams or students taking the seventh-grade mathematics and literacy Benchmark Exams. The sixth- and seventh-grade classes at each of the school districts ranged between 47% and 51% female and from 39 % to 70 % Caucasian. All students in the study were tested in literacy and mathematics on the Arkansas Benchmark Exam for the 2012-2013 school year. All school districts were within a 50-mile radius of one another.

Demographics

All students received instruction for literacy and mathematics and were designated to either the sixth or seventh grade. In this study, 1,441 sixth graders used TLI interim assessments. Additionally, 2,189 sixth graders used NWEA interim assessment. 1,331 seventh graders used TLI interim assessment. Further, 2,192 seventh graders used NWEA interim assessment. Of the sixth-grade students using TLI, 50% to 51% were female and 54% to 63% were Caucasian. Of the sixth-grade students using NWEA, 47% to 48% were female and 40% to 69% were Caucasian. Of the seventh-grade students using TLI, 50% were female and 56% to 67% were Caucasian. Of the seventh-grade

students using NWEA, 48% to 49% were females and 39% to 69% were Caucasian. All four school districts used are located in Northwest Arkansas. The sample and demographics explained above represent the accessible populations. For this study, a stratified random sample was executed from these available populations.

Instrumentation

The Arkansas Comprehensive Testing, Assessment, and Accountability Program Arkansas Benchmark Exams for sixth- and seventh-grade literacy and mathematics were used for this study. The augmented Benchmark Exam is a criterion-referenced test. It includes multiple-choice and open-response questions designed to assess students' proficiency in Arkansas' learning standards in literacy and mathematics.

The format of the sixth- and seventh-grade Benchmark Exams is identical. Each session of the mathematics' exams contains 20 multiple-choice questions. However, there are either 2 or 3 open-response questions for each session of the mathematics' exams. Specifically, there are 40 multiple-choice questions on the exams with strand topics consisting of number and operations, algebra, geometry, measurement, and data analysis and probability. There is 1 open-response question from each of those strands for a total of 5.

For the literacy Benchmark Exams, each session contains 16 multiple-choice questions. However, the sessions are broken down into two types of formats. The first type consists of 16 multiple-choice questions for reading with 2 open-response questions for reading and an essay for writing. The second type of format for the literacy sessions consists of 8 multiple-choice questions for reading and 8 multiple-choice questions for writing. Additionally, this session format includes an open-response question for reading

and an essay for writing. Specifically, there are 32 total multiple-choice questions with strand topics consisting of literary passage, content passage, practical passage, and writing. There are 13 open-response items with 3 connected with the reading strands and 10 connected with the writing strands.

Each of the Arkansas Benchmark Exams used in the study were tested for reliability. The 2013 Arkansas Augmented Benchmark Examinations Technical Report (Questar Assessment, 2013) stated the reliability (stratified alpha) for the Grade 6 Mathematics Exam is 0.89. The reliability (stratified alpha) for the Grade 7 Mathematics Exam is 0.89. The reliability (stratified alpha) for the Grade 6 Literacy Exam is 0.89. The reliability (stratified alpha) for the Grade 7 Literacy Exam is 0.88.

Data Collection

Each selected district collected data without personal identification information. These data were requested from the school districts that participated in either TLI or NWEA for the 2012-2013 school year. The data set included the gender of the student, grade placement of the student, literacy raw score of the student, and mathematics raw score of the student. Summarizing, every sixth- or seventh-grade student was categorized with a male or female designation and a score for both literacy and mathematics. The data were received by email during the summer and fall semesters of 2014 and the spring semester of 2015. A stratified random sample was implemented for equal-sized samples. The data were stored on a USB flash drive and locked in a fireproof safe when not being used.

Analytical Methods

For the purpose of data analysis, *IBM Statistical Packages for Social Sciences Version 22* was used. After collecting data, the student's grade, gender, raw literacy score and raw mathematics score were coded in the system. Each of the four hypotheses was analyzed using the following statistical analysis. A pre-analysis was conducted to verify the number of participants by grade level, gender, raw literacy score, and raw mathematics score. A second analysis was conducted to identify outliers. Further, Levene's test was used to check homogeneity of variances.

To address the first hypothesis, a 2 x 2 factorial ANOVA was conducted using the type of interim assessment and gender as the independent variables and literacy achievement of sixth-grade students as the dependent variable. To address the second hypothesis, a 2 x 2 factorial ANOVA was conducted using the type of interim assessment and gender as the independent variables and literacy achievement of seventh-grade students as the dependent variable. To address the third hypothesis, a 2 x 2 factorial ANOVA was conducted using the type of interim assessment and gender as the independent variables and mathematics achievement of sixth-grade students as the dependent variable. To address the fourth and final hypothesis, a 2 x 2 factorial ANOVA using the type of interim assessment and gender as the independent variables and mathematics achievement of seventh-grade students as the dependent variable. For all the tests in this study, a two tailed test with a .05 level was used.

Limitations

Limitations of a study need to be noted in order to enhance the reader's understanding in how to interpret its results. A causal-comparative study has a built in

limitation because the results of the study occurred before the study takes place. It is not an experimental study. Thus, the independent variables could not be manipulated or participants randomly assigned. However, this limitation did not appear to inhibit the study using schools for research purposes. Next, the sample only included sixth- and seventh-grade students in four school districts located in Northwest Arkansas. Because of these restrictions, the procedures as well as the results of this study cannot be applied to all schools and situations. Certainly, researching younger or older students or researching private schools or public schools in a different area of the country may yield different results.

Third, with the volatile climate in curriculum and assessment in the state of Arkansas and around the country, using Benchmark Exam results might have a limited usefulness in determining a program's effectiveness. From the 2013-2014 school year to the current school year, Arkansas has used the Arkansas Comprehensive Testing, Assessment, and Accountability Program, the Partnership for Assessment of Readiness of College and Career, and now the ACT-Aspire in successive years. Next, students observed in the study had all taken Arkansas Comprehensive Testing, Assessment, and Accountability Program and other standardized tests and might have recognized certain test questions from this experience. This potential limitation may have affected internal validity. However, the length of time between tests would seem to ensure that this was not a major limitation.

Fourth, though all school districts used in the sample are within a 50-mile radius, there are some demographic differences in the districts. However, the students using TLI and NWEA seemed to balance out some of the differences proportionally. Regardless of

limitations, considering the use of TLI or NWEA as an interim assessment program to prepare students for Benchmark Exams provides research that is scarce and needed. School districts spend large amounts of money to purchase an interim assessment program with the intention of it improving student performance. This study provided information that might impact this decision making.

CHAPTER IV

RESULTS

The purpose of the study was four-fold. A quantitative approach was adopted in this study to examine the effect of type of interim assessment by gender on literacy and mathematics performance from a sample of sixth-grade students in Northwest Arkansas. In addition, this study examined the effect of type of assessment by gender on literacy and mathematics performance from a sample of seventh-grade students in Northwest Arkansas. The independent variables for all the hypotheses were the type of interim assessment and gender. The dependent variables were literacy and mathematics performance. A factorial ANOVA was used to test each hypothesis. The IBM Statistical Packages for Social Sciences software was used to run assumptions of homogeneity of variances, analyze independence of groupings, and check for outliers. Results of this analysis are found in this chapter.

Hypothesis 1

Hypothesis 1 stated that no significant differences will exist by gender in Benchmark Literacy Exam performance of sixth-grade students in Arkansas middle schools who received TLI-interim assessment versus those sixth-grade students in Arkansas middle schools who received NWEA-interim assessment. Data were screened for data entry errors and missing values. No data entry errors were found. One NWEA male student had no score for literacy. The Kolmogorov-Smirnov (KS) statistics and

histograms were used to test the assumption of normality. An examination of the histogram for each group revealed some deviations from a normal distribution. Results from the KS tests confirmed these violations for the literacy performance distribution of NWEA males, $D(99) = 0.13, p = .000$, TLI males, $D(100) = 0.14, p = .000$, as well as for NWEA females, $D(100) = 0.14, p = .000$. However, KS results showed no significant skew/kurtosis in the distribution of literacy performance of TLI females, $D(100) = 0.08, p > .05$. Despite the observed violations in normality, proceeding with the analysis of data using ANOVA was deemed appropriate because ANOVA is considered robust to violations of the assumption of normality (Pallant, 2013). Table 1 presents a summary the group means and standard deviations for this analysis.

Table 1

Descriptive Statistics from Sixth-Grade 2013 Arkansas Benchmark Literacy Exam Raw Scores

Assessment	Sixth-grade literacy							
	Male			Female			Total	
	<i>N</i>	<i>M</i>	<i>SD</i>	<i>N</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
NWEA	99	666.62	183.19	100	730.30	135.56	698.62	163.76
TLI	100	701.45	175.03	100	774.05	113.72	737.75	151.66
Total	199	684.12	179.54	200	752.18	126.72	718.23	158.82

Next, to test the assumption of equality of variances, Levene’s test was conducted within ANOVA and indicated that the assumption of variances was violated, $F(3,395) = 7.80, p = .000$. However, the ANOVA test is reasonably robust to violations of this

assumption when group size is similar (Pallant, 2013). Further, examination of the data revealed no significant outliers. Having checked all the assumptions associated with ANOVA, Hypothesis 1 was tested using a 2 x 2 factorial ANOVA to evaluate the effects of interim assessment type and gender on literacy achievement as measured by 2013 Arkansas Benchmark Literacy Exam. Results of this analysis are displayed in Table 2.

Table 2

Factorial ANOVA Results from 2013 Arkansas Sixth-Grade Benchmark Literacy Exam Raw Scores

Source	<i>SS</i>	<i>df</i>	<i>MS</i>	<i>F</i>	<i>p</i>	<i>ES</i>
Gender	463162.51	1	463162.51	19.42	.000	0.05
Assessment	153996.61	1	153996.61	6.46	.011	0.02
Gender*Assessment	1982.44	1	1982.44	0.08	.773	0.00
Error	9421331.91	395	23851.47			

These results revealed no significant interaction between assessment and gender on the literacy performance of sixth-grade students, $F(1,395) = 0.08, p = .773, ES = 0.00$. Given there was no significant interaction between the variables of gender and assessment, the main effect of each variable was examined separately. The main effect for gender was significant with a small effect size, $F(1,395) = 19.42, p = .000, ES = 0.05$. The main effect of assessment was also significant with a small effect size, $F(1,395) = 6.46, p = .011, ES = 0.02$. Figure 1 shows the means for sixth-grade literacy achievement as a function of type of assessment and gender.

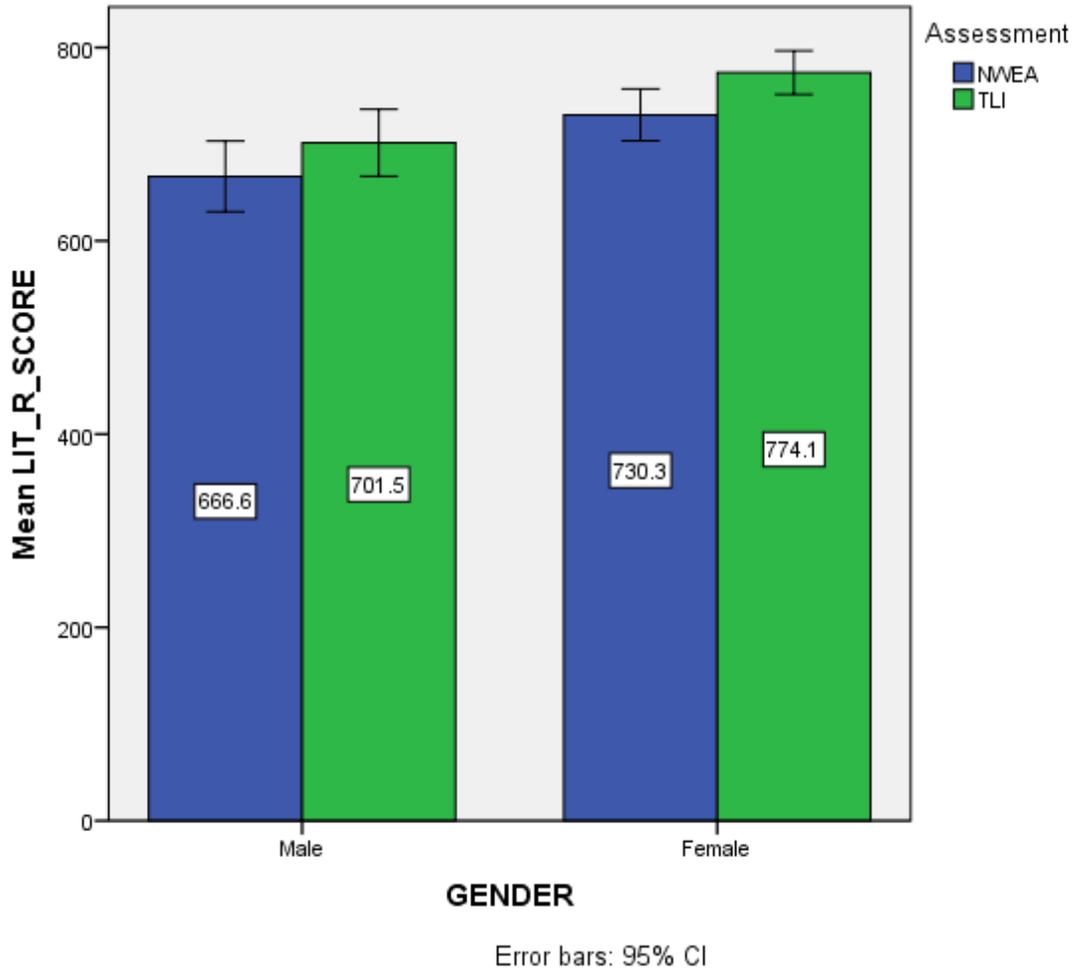


Figure 1. Mean sixth-grade Benchmark Literacy scores by gender and assessment.

The results show that females outperformed males, regardless of the type of assessment, and TLI students outperformed NWEA students, regardless of gender. On the basis of these results, the null hypothesis could not be rejected for the interaction between gender and assessment. However, the main effect hypotheses for both assessment and gender were rejected.

Hypothesis 2

Null hypothesis 2 stated that no significant differences will exist by gender in Benchmark Literacy Exam performance of those seventh-grade students in Arkansas

middle schools who received TLI-interim assessment versus those seventh-grade students in Arkansas middle schools who received NWEA-interim assessment. Data were screened for data entry errors and missing values. No cases of data entry errors or missing values were found. The KS statistics and histograms were used to test the assumption of normality. An examination of the histogram for each group revealed some deviations from a normal distribution. Results from the KS tests confirmed these violations for the literacy performance distribution of NWEA males, $D(100) = 0.12, p = .001$, TLI males, $D(100) = 0.11, p = .005$, as well as for NWEA females, $D(100) = 0.10, p = .010$, and TLI females, $D(100) = 0.13, p = .000$. Despite the observed violations in normality, proceeding with the analysis of data using ANOVA was deemed appropriate because ANOVA is considered robust to violations of the assumption of normality (Pallant, 2013). Table 3 presents a summary the group means and standard deviations for this analysis.

Table 3

Descriptive Statistics from Seventh-Grade 2013 Arkansas Benchmark Literacy Exam Raw Scores

Assessment	Seventh-grade literacy							
	Male			Female			Total	
	<i>N</i>	<i>M</i>	<i>SD</i>	<i>N</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
NWEA	100	693.95	155.45	100	746.15	139.32	720.05	149.54
TLI	100	734.65	133.43	100	789.20	120.96	761.93	129.93
Total	200	714.30	145.93	200	767.68	131.91	740.99	141.47

Next, to test the assumption of equality of variances, Levene’s test was conducted within ANOVA and indicated that the assumption of variances was not violated, $F(3,396) = 1.64, p > .05$. Further examination of the data revealed no significant outliers. Having checked all the assumptions associated with ANOVA, Hypothesis 2 was tested using a 2 x 2 factorial ANOVA to evaluate the effects of interim assessment type and gender on literacy achievement as measured by 2013 Arkansas Benchmark Literacy Exam for the seventh-grade students. Results of this analysis are displayed in Table 4.

Table 4

Factorial ANOVA Results from 2013 Arkansas Seventh-Grade Benchmark Literacy Exam Raw Scores

Source	<i>SS</i>	<i>df</i>	<i>MS</i>	<i>F</i>	<i>p</i>	<i>ES</i>
Gender	284889.06	1	284889.06	14.99	.000	0.04
Assessment	175351.56	1	175351.56	9.23	.003	0.02
Gender*Assessment	138.06	1	138.06	0.01	.932	0.00
Error	7524706.25	396	19001.78			

These results revealed no significant interaction between assessment and gender on the literacy performance of seventh-grade students, $F(1,396) = 0.01, p = .932, ES = 0.00$. Given there was no significant interaction between the variables of gender and assessment, the main effect of each variable was examined separately. The main effect for gender was significant with a small effect size, $F(1,396) = 14.99, p = .000, ES = 0.04$. The main effect of assessment was also significant with a small effect size, $F(1,396) =$

9.23, $p = .003$, $ES = 0.02$. Figure 2 shows the means for seventh-grade literacy achievement as a function of type of assessment and gender.

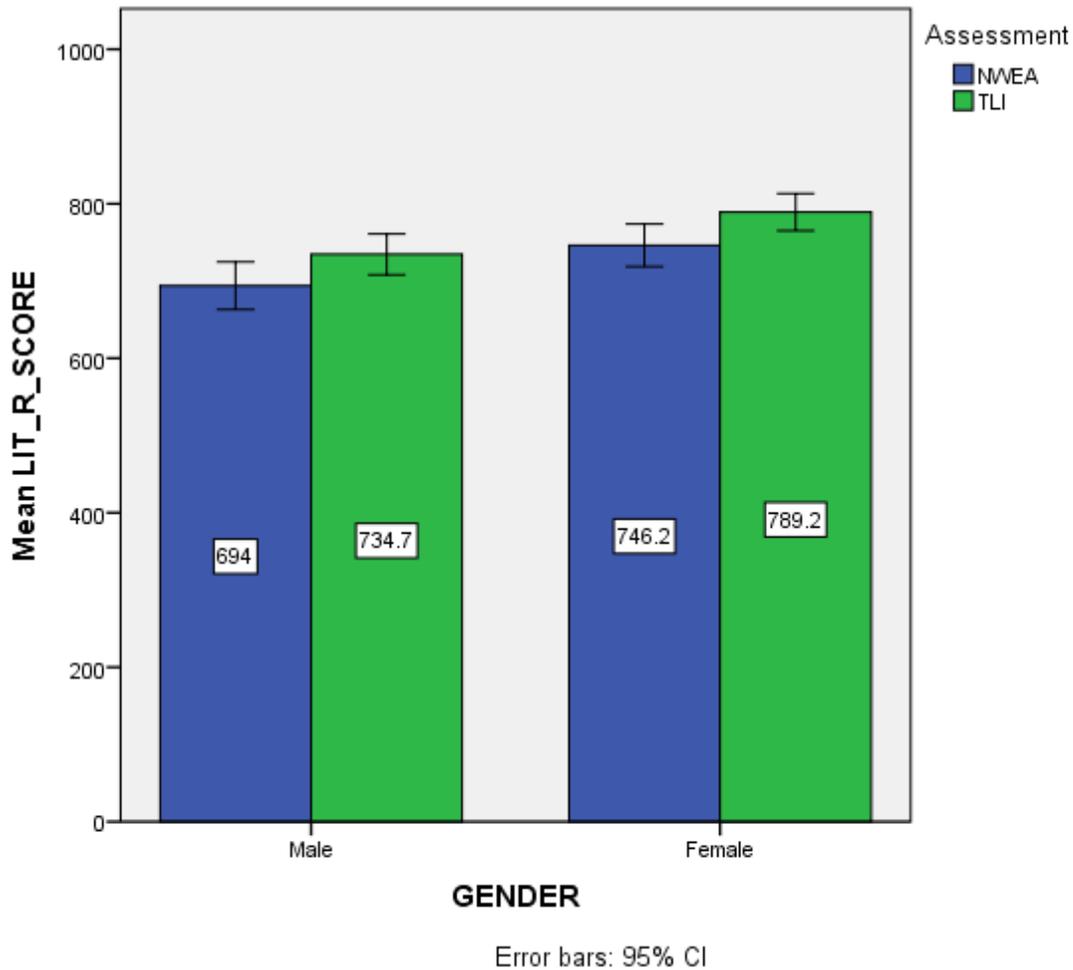


Figure 2. Mean seventh-grade Benchmark Literacy scores by gender and assessment.

The results showed that females outperformed males, regardless of the type of assessment, and TLI students outperformed NWEA students, regardless of gender. On the basis of these results, the null hypothesis could not be rejected for the interaction between gender and assessment. However, the main effect hypotheses for both assessment and gender were rejected.

Hypothesis 3

Null hypothesis 3 stated that no significant differences will exist by gender in Benchmark Mathematics Exam performance of sixth-grade students in Arkansas middle schools who received TLI-interim assessment versus those sixth-grade students in Arkansas middle schools who received NWEA-interim assessment. Data were screened for data entry errors and missing values. No cases of data entry errors or missing values were found. The KS statistics and histograms were used to test the assumption of normality. An examination of the histogram for each group revealed some deviations from a normal distribution. Results from the KS tests confirmed these violations for the mathematics performance distribution of NWEA males, $D(100) = 0.11, p = .003$. However, results from the KS tests revealed no violation for the mathematics performance distribution of TLI males, $D(100) = 0.07, p > .05$, as well as for NWEA females, $D(100) = 0.09, p > .05$, and TLI females, $D(100) = 0.06, p > .05$. Despite the observed violations in normality, proceeding with the analysis of data using ANOVA was deemed appropriate because ANOVA is considered robust to violations of the assumption of normality (Pallant, 2013). Table 5 presents a summary of group means and standard deviations for this analysis.

Table 5

Descriptive Statistics from Sixth-Grade 2013 Arkansas Benchmark Mathematics Exam Raw Scores

Assessment	Sixth-grade mathematics							
	Male			Female			Total	
	<i>N</i>	<i>M</i>	<i>SD</i>	<i>N</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
NWEA	100	441.20	167.95	100	456.30	152.27	448.75	160.08
TLI	100	433.80	154.61	100	465.80	134.59	449.80	145.47
Total	200	437.50	161.05	200	461.1	143.42	449.28	152.76

Next, to test the assumption of equality of variances, Levene's test was conducted within ANOVA and indicated that the assumption of variances was violated, $F(3,396) = 1.64, p = .036$. However, the ANOVA test is reasonably robust to violations of this assumption when group size is similar (Pallant, 2013). Further examination of the data revealed no significant outliers. Having checked all the assumptions associated with ANOVA, Hypothesis 3 was tested using a 2 x 2 factorial ANOVA to evaluate the effects of interim assessment type and gender on mathematics achievement as measured by 2013 Arkansas Benchmark Mathematics Exam. Results of this analysis are displayed in Table 6.

Table 6

Factorial ANOVA Results from 2013 Arkansas Sixth-Grade Benchmark Mathematics Exam Raw Scores

Source	<i>SS</i>	<i>df</i>	<i>MS</i>	<i>F</i>	<i>p</i>	<i>ES</i>
Gender	55460.25	1	55460.25	2.38	.124	0.01
Assessment	110.25	1	110.25	0.01	.945	0.00
Gender*Assessment	7140.25	1	7140.25	0.31	.581	0.00
Error	9247779.00	396	23353.00			

These results revealed no significant interaction between assessment and gender on the mathematics performance of sixth-grade students, $F(1,396) = 0.31, p = .581, ES = 0.00$. Given there was no significant interaction between the variables of gender and assessment, the main effect of each variable was examined separately. The main effect for gender was not significant with a small effect size, $F(1,396) = 2.38, p > 0.05, ES = 0.01$. The main effect of assessment was also not significant with a small effect size, $F(1,396) = 0.01, p > 0.05, ES = 0.00$. Figure 3 shows the means for sixth-grade mathematics achievement as a function of type of assessment and gender.

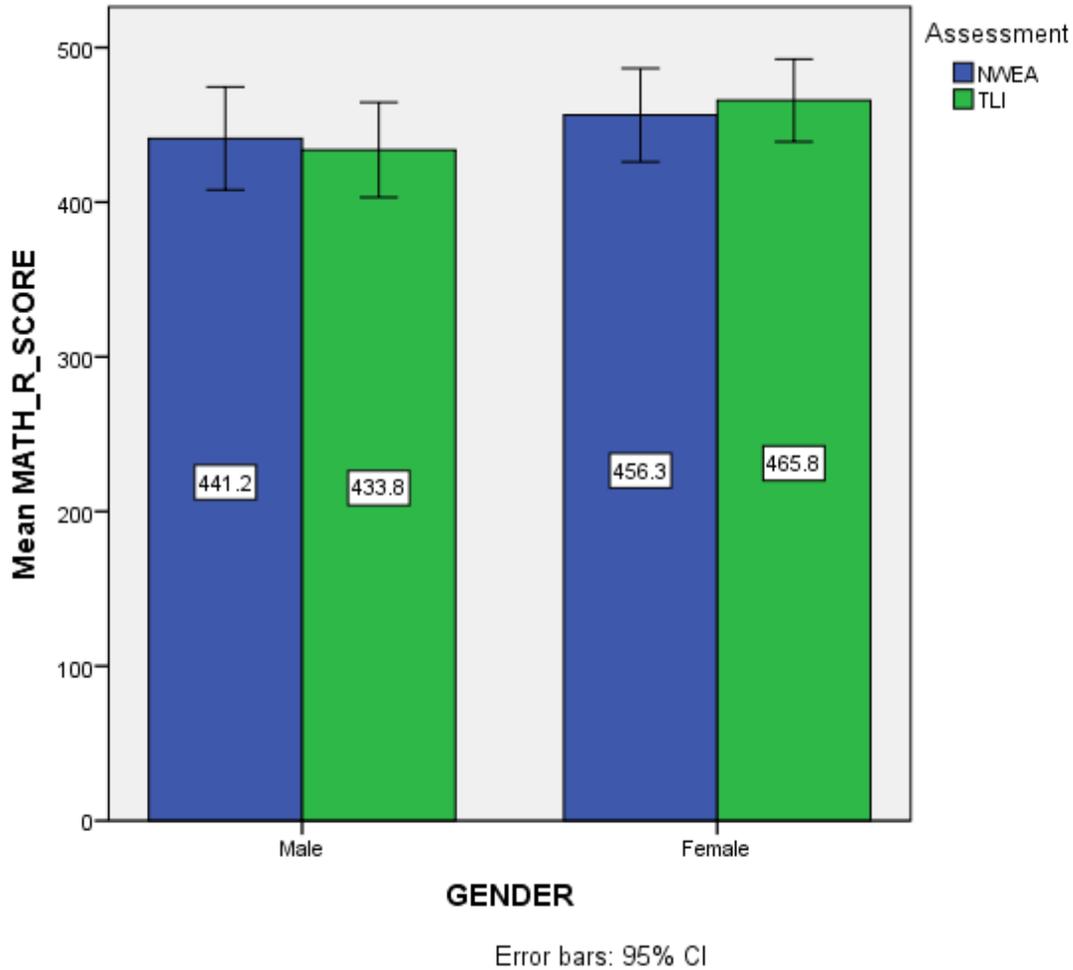


Figure 3. Mean sixth-grade Benchmark Mathematics scores by gender and assessment.

The results showed that, even though females performed higher compared to males, their performance was not significantly higher. Similarly, even though TLI participants scored, on average, higher compared to the NWEA participants, the difference was not significant. On the basis of these results, the null hypothesis could not be rejected for the interaction between gender and assessment. Furthermore, the main effect hypotheses for both assessment and gender were not rejected.

Hypothesis 4

Null hypothesis 4 stated that no significant differences will exist by gender in Benchmark Mathematics Exam performance of seventh-grade students in Arkansas

middle schools who received TLI-interim assessment versus those seventh-grade students who received NWEA-interim assessment. Data were screened for data entry errors and missing values. No cases of data entry errors or missing values were found. The KS statistics and histograms were used to test the assumption of normality. An examination of the histogram for each group revealed some deviations from a normal distribution. Results from the KS tests confirmed this violation for the mathematics performance distribution of TLI males, $D(100) = 0.13, p = .000$. However, results from the KS tests revealed no violation for the mathematics performance distribution of NWEA males, $D(100) = 0.06, p > .05$, as well as for NWEA females, $D(100) = 0.09, p > .05$, and TLI females, $D(100) = 0.05, p > .05$. Despite the observed violations in normality, proceeding with the analysis of data using ANOVA was deemed appropriate because ANOVA is considered robust to violations of the assumption of normality (Pallant, 2013). Table 7 presents a summary the group means and standard deviations for this analysis.

Table 7

Descriptive Statistics from Seventh-Grade 2013 Arkansas Benchmark Mathematics Exam Raw Scores

Assessment	Seventh-grade mathematics							
	Male			Female			Total	
	<i>N</i>	<i>M</i>	<i>SD</i>	<i>N</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
NWEA	100	400.90	154.79	100	396.50	141.07	398.70	147.73
TLI	100	399.90	165.65	100	414.30	142.90	407.10	154.47
Total	200	400.40	159.91	200	405.40	141.91	402.90	151.01

Next, to test the assumption of equality of variances, Levene’s test was conducted within ANOVA and indicated that the assumption of variances was violated, $F(3,396) = 2.69, p = .046$. However, the ANOVA test is reasonably robust to violations of this assumption when group size is similar (Pallant, 2013). Further examination of the data revealed no significant outliers. Having checked all the assumptions associated with ANOVA, Hypothesis 4 was tested using a 2 x 2 factorial ANOVA to evaluate the effects of interim assessment type and gender on mathematics achievement as measured by 2013 Arkansas Benchmark Mathematics Exam. Results of this analysis are displayed in Table 8.

Table 8

Factorial ANOVA Results from 2013 Arkansas Seventh-Grade Benchmark Mathematics Exam Raw Scores

Source	<i>SS</i>	<i>df</i>	<i>MS</i>	<i>F</i>	<i>p</i>	<i>ES</i>
Gender	2500.00	1	2500.00	0.11	.741	0.00
Assessment	7056.00	1	7056.00	0.31	.579	0.00
Gender*Assessment	8836.00	1	8836.00	0.39	.535	0.00
Error	9080244.00	396	22929.91			

These results revealed no significant interaction between assessment and gender on the mathematics performance of seventh-grade students, $F(1,396) = 0.39, p = .535, ES = 0.00$. Given there was no significant interaction between the variables of gender and assessment, the main effect of each variable was examined separately. The main effect for gender was not significant with a small effect size, $F(1,396) = 0.11, p > .05, ES =$

0.00. The main effect of assessment was also not significant with a small effect size, $F(1,396) = 0.31, p > .05, ES = 0.00$. Figure 4 shows the means for seventh-grade mathematics achievement as a function of type of assessment and gender.

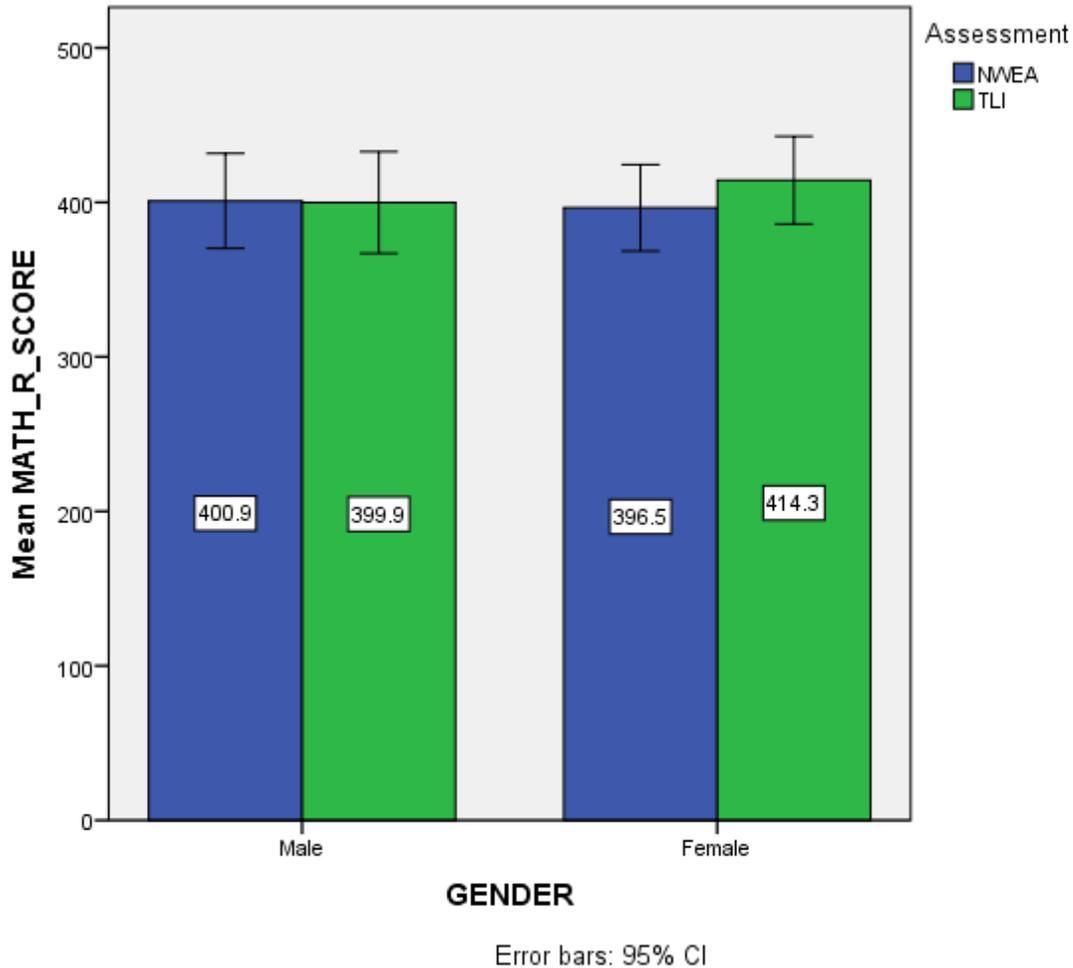


Figure 4. Mean seventh-grade Benchmark Mathematics scores by gender and assessment.

The results showed that males and females performed similarly, regardless of the type of assessment, and TLI students and NWEA students performed similarly, regardless of gender. On the basis of these results, the null hypothesis could not be rejected for the interaction between gender and assessment. Furthermore, the main effect hypotheses for both assessment and gender or assessment were also not rejected.

Summary

The purpose of this study was to determine the effects of gender and assessment type on literacy and mathematics achievement for sixth and seventh-grade students in Northwest Arkansas. For the four hypotheses, none displayed a significant interaction between assessment type and gender. See Table 9 for results of significance for interaction and main effect of variables.

Table 9

Summary of Statistically Significant Results ($p \leq .05$) by Grade for Hypotheses 1-4,

Variables	Literacy p		Mathematics p	
	H1	H2	H3	H4
Gender*Assessment	.773	.932	.581	.535
Assessment	.011*	.003**	.945	.579
Gender	.000**	.000**	.124	.741
Grade				
6	✓		✓	
7		✓		✓

* $p < .05$

** $p < .01$

Additionally, neither the main effect for gender nor the main effect for assessment type was not significant for the two hypotheses related to mathematics achievement for sixth or seventh-grade students. However, the two hypotheses related to literacy achievement yielded different results. The main effect of both gender and assessment type were significant in regard to literacy achievement. The effect size of assessment was

small for both sixth- and seventh-grade students on literacy achievement. The effect size for gender was also small for both sixth- and seventh-grade students on literacy achievement.

CHAPTER V

DISCUSSION

The effects of interim assessment on academic achievement represent an important factor to consider for current school administrators. Fleming (2013) emphasized the importance of recognizing each student's current instructional level to assess and meet each student's academic needs. Interim assessment programs may provide student achievement data and measure the different levels of performance. The insight from this information might provide the benefit of improved instruction. However, Goertz, Olah, and Riggan (2009) emphasized that although interim assessments were useful, they were not sufficient standing alone to address improved instruction. The data gained from interim assessment did inform teachers over what content should be re-taught to students. Interim assessments were aligned to state standards and curriculum which helped school and district personnel to use the resulting data better. Interestingly, school leadership and culture of data were identified as critical factors for improving instruction as well as interim assessments. Effective instruction is an important aspect to consider when focusing on improved academic achievement. Interim assessment's impact on instruction was not clear from the research.

Every culture in public schools across the country is unique to specific school and community. Tailoring the interim assessment program to the needs of the students in a school district is important. Ryan (2010) stated that attention should focus on the

program's validity for all students aligning resources, finances, material, and personnel to meet those needs. An effective evaluation system would require constant monitoring, reviewing, and adjusting to ensure the program would be successful to justify the district's investment. Further, Hicks (2014) stated that the purchaser must consider the adaptiveness of the assessment. For the interim assessment program to meet its intended needs, the data must be reliable to influence building and district level decisions. Lastly, the data should help to predict high stakes test outcomes like the benchmark exams that schools are considering.

For this study, the Benchmark results in literacy and mathematics and the effects on these tests by interim assessment programs (NWEA versus TLI) and by gender were examined. Findings indicated that TLI costs significantly more than NWEA. Students who received TLI interim assessment preparation scored significantly higher than students that received NWEA interim assessment preparation on the literacy Benchmark Exam for both sixth- and seventh-grades. Further, females did perform better in literacy than their male counterparts. There was no significant difference in achievement in mathematics by interim assessment programs or by gender. This chapter includes conclusions for the hypotheses, implications from the study, and recommendations regarding the potential for practice and policy and for future research considerations.

Conclusions

To address the four hypotheses, the following statistical analyses were used. Hypothesis 1 was analyzed using a 2 x 2 factorial ANOVA with gender and assessment type as the between-subjects independent variables with sixth-grade literacy Benchmark Exam raw scores as the dependent variable. Hypothesis 2 was analyzed using a 2 x 2

factorial ANOVA with gender and assessment type as the between-subjects independent variables with seventh-grade literacy Benchmark Exam raw scores as the dependent variable. Hypothesis 3 was analyzed using a 2 x 2 factorial ANOVA with gender and assessment type as the between-subjects independent variables with sixth-grade mathematics Benchmark Exam raw scores as the dependent variable. Hypothesis 4 was analyzed using a 2 x 2 factorial ANOVA with gender and assessment type as the between-subjects independent variables with seventh-grade mathematics Benchmark Exam raw scores as the dependent variable. To test the null hypotheses, a two-tailed test with a .05 level of significance was used. Interaction and main effects were examined for each of the hypotheses. The following hypotheses were tested and used to determine conclusions.

Hypothesis 1

Hypothesis 1 stated that no significant differences will exist by gender in Benchmark Literacy Exam performance of sixth-grade students in Arkansas middle schools who received TLI-interim assessment versus those sixth-grade students in Arkansas middle schools who received NWEA-interim assessment. In analyzing the results of the ANOVA, the type of assessment used did not interact with gender on students' literacy achievement. Therefore, no significant interaction effect was noted between assessment and gender on the literacy test performance of sixth-grade students, and the null hypothesis for the interaction effect was not rejected. However, the main effect of assessment was statistically significant. Students using the TLI assessment performed better on the Benchmark Literacy Exam compared to students using the NWEA assessment regardless of their gender. Additionally, the main effect of gender

was also statistically significant. In general, females outperformed males on the literacy Benchmark Exam for sixth-grade students regardless of the type of assessment.

Consequently, the main effects for both assessment type and gender were significant, and the null hypotheses for both main effects were rejected.

Hypothesis 2

Hypothesis 2 stated no significant differences will exist by gender in Benchmark Literacy Exam performance of those seventh-grade students in Arkansas middle schools who received TLI-interim assessment versus those seventh-grade students in Arkansas middle schools who received NWEA-interim assessment. In analyzing the results of the ANOVA, the type of assessment used did not interact with gender on students' literacy achievement. Therefore, no significant interaction effect was noted between assessment and gender on the literacy test performance of seventh-grade students, and the null hypothesis for the interaction effect was not rejected. However, the main effect of assessment was statistically significant. Students using the TLI assessment performed better on the Benchmark Literacy Exam compared to students using the NWEA assessment regardless of their gender. Additionally, the main effect of gender was also statistically significant. In general, females outperformed males on the literacy Benchmark Exam for seventh-grade students regardless of the type of assessment. Consequently, the main effects for both assessment type and gender were significant, and the null hypotheses for both main effects were rejected.

Hypothesis 3

Hypothesis 3 stated no significant differences will exist by gender in Benchmark Mathematics Exam performance of sixth-grade students in Arkansas middle schools who

received TLI-interim assessment versus those sixth-grade students in Arkansas middle schools who received NWEA-interim assessment. In analyzing the results of the ANOVA, the type of assessment used did not interact with gender on students' mathematics achievement. Therefore, no significant interaction effect was noted between assessment and gender on the mathematics test performance of sixth-grade students, and the null hypothesis for the interaction effect was not rejected. Likewise, the main effect of assessment was not statistically significant. Students using the TLI assessment performed no differently on the Benchmark Mathematics Exam compared to students using the NWEA assessment regardless of their gender. Additionally, the main effect of gender was also not statistically significant. In general, females performed similarly to males on the mathematics Benchmark Exam for sixth-grade students regardless of the type of assessment. Consequently, the main effects for both assessment type and gender were not significant, and the null hypotheses for both main effects were not rejected.

Hypothesis 4

Hypothesis 4 stated no significant differences will exist by gender in Benchmark Mathematics Exam performance of seventh-grade students in Arkansas middle schools who received TLI-interim assessment versus those seventh-grade students who received NWEA-interim assessment. In analyzing the results of the ANOVA, the type of assessment used did not interact with gender on students' mathematics achievement. Therefore, no significant interaction effect was noted between assessment and gender on the mathematics test performance of seventh-grade students, and the null hypothesis for the interaction effect was not rejected. Likewise, the main effect of assessment was not statistically significant. Students using the TLI assessment performed no differently on

the Benchmark Mathematics Exam compared to students using the NWEA assessment regardless of their gender. Additionally, the main effect of gender was also not statistically significant. In general, females performed similarly to males on the mathematics Benchmark Exam for seventh-grade students regardless of the type of assessment. Consequently, the main effects for both assessment type and gender were not significant, and the null hypotheses for both main effects were not rejected.

Implications

To interpret the results of this study, a reflection over the larger context of the review of literature should occur. Research regarding the influence of interim assessments on student achievement has yielded mixed results. For instance, Blazer (2008) stated that preliminary indicators of a study indicated that an interim assessment program had a positive effect on student achievement. However, she indicated that the predictive validity of interim assessments produced mixed findings and a more detailed study in the future would be needed. McGraw-Hill Education (2015) reviewed an interim assessment program's effect on reading and mathematics. The results of the study showed a statistically positive impact on literacy in Grades 3 through 5 and on mathematics in Grades 5 and 6. Several studies were similar, finding varying results depending on the age group and subject. For example, Konstantopoulos et al. (2013) stated the vast majority of the treatment effects in their Indiana study regarding the effects of interim assessment on student achievement for elementary students were positive but not significant at most levels. The treatment effect was significant, however, in fifth- and sixth-grade students for mathematics achievement.

While the aforementioned research relates to this study, clear distinctions in the findings are evident and not directly applicable. The study in this project was not comparing the use of interim assessments in comparison to traditional instruction or previously used instructional methods. The study focused on the comparison of two rival interim assessment providers effect on student achievement for mathematics and literacy. Research that compares two or more specific providers of interim assessments was not found. Therefore, this study should provide some new information for comparing interim assessments. For this study, TLI clearly outperformed NWEA at a significant level in sixth- and seventh-grade literacy. There was no significant difference in the two interim assessments in mathematics performance of sixth- and seventh-grade students. Although the cost of TLI is significantly more than NWEA, the difference in literacy scores was at a significant level. A school district's financial resources might be the determining factor in whether or not the district could pursue purchasing the interim assessment program.

As mentioned earlier in the study, there are many rules of thoughts or even myths regarding the effect of gender on academic achievement. The thoughts that girls are better in English and boys are better at math are not consistently supported by research. However, some research exists to support such notions. Zembar and Blume (2011) stated that, while there was no difference in girls and boys mathematics performance in fourth grade, girls tended to lose ground in mathematics performance beyond fourth grade. Further, girls tend to excel in verbal ability over boys throughout high school. Klinger et al. (2009) reiterated girls' superiority in reading and writing stating that girls outperforming boys had been a static pattern over the past 40 years. However, they did report some cases of boys performing similarly to girls in literacy. Moreover, they stated

that the historical gap of boys outperforming girls in mathematics has shrunk in many instances.

Furthermore, Hyde and Mertz (2009) stated: "...studies find more males than females scoring above the 95th or 99th percentile, but the gender gap has narrowed over time in the U.S." (p. 1). The authors also emphasized that the gaps in mathematics are not found at all in some ethnic groups and some countries. Last, Penner and Paret (2008) stated that gender differences were thought to emerge in mathematics at the end of middle school and beginning of high school with boys outperforming girls. However, they stated that the change occurs as early as third grade. As stated, some research supports girls' dominance in literacy and boys' dominance in mathematics, but exceptions are revealed as well.

For this study, girls significantly outperformed boys in literacy achievement for the Arkansas Benchmark Exams for both sixth- and seventh-graders. The girls scored higher regardless of what interim assessment program was used. The results of the study were consistent with much of the research concerning gender effects on literacy achievement. However, gender effects on mathematics performance did not follow with the research that boys tend to outperform girls in mathematics. In fact, there was no significant difference in mathematics performance of the Arkansas Benchmark Exams between boys and girls for sixth and seventh-graders. In fact, the mean scores for girls were higher than the mean scores for boys in mathematics performance for both sixth- and seventh-grade students though not at a significant level.

Recommendations

Potential for Practice/Policy

The study was conducted in four school districts in Northwest Arkansas and limited to sixth- and seventh-grade students only. The study compared the test scores of sixth-grade students in literacy and mathematics who received either TLI or NWEA interim assessments, as well as seventh-grade student in the same subject areas. The study compared the test scores by gender in those grade levels, as well. The findings of the study might have an impact on the policies for purchasing of interim assessments by administrations in school districts. First, school districts should determine if interim assessment programs are worth the expense. The cost of an interim assessment program might be a major factor to consider for some school districts. Some school districts generate more unrestricted local wealth than other school districts. However, some poorer school districts might receive additional categorical revenue in which educational programs may be purchased. A school district would need to conduct a cost analysis evaluating the cost of the product versus the research that identifies the strengths of the interim assessment program and equate the fiscal resources to purchase with the academic needs of the district. All school districts are different as are their specific needs for student learning.

Second, school districts should research whether some interim assessment programs are better than others. For instance, some interim assessments might perform better in literacy and other programs perform better in mathematics. Educators must research which program best suits their school. Do some interim assessment programs reach the needs of girls better than boys? For example, a school district might have a

scenario where the importance of education is more important for boys than girls or vice versa. If an interim assessment program yields data that it can be effective with the gender the district needs help with, then that interim assessment program might be the most appropriate. Some districts create their interim assessments. Some interim assessment programs are inexpensive or free. A district's school leaders must determine what is better for their school.

Third, the state legislature may determine that it is necessary to pass laws so that all school districts can afford interim assessments. Some districts create their own assessments, and some are available at little or no cost. In theory, any district may obtain an interim assessment program. However, some districts may not be able to purchase the interim assessment program that their students need because of cost. The discrepancy in funding for individual districts may be considered an equity issue. It is important that every student in the United States have an education that meets their individual needs. State comprehensive exams determine how public schools are rated. ACT and SAT scores are factors in determining students' funding for higher education. Therefore, testing is important in today's educational system. Preparing students for these assessment challenges must be at the forefront of the concerns for government officials.

Fourth, school officials should determine how to address gender gaps in mathematics and literacy. While research yields mixed results on gender differences in student achievement, enough research supports certain discrepancies in academic achievement in boys and girls that it should be considered. Programs should be researched that might improve girls' achievement in mathematics and boys' achievement in literacy. Interim assessment programs might be helpful to balance the scales, but other

options should be considered as well. Programs that address the stereotyping that could affect a student's mindset on what he or she should or should not enjoy at school should be researched. Determining what differences are due to lack of academic potential in an area or other contributing factors should be studied.

Fifth, the school district should provide research-based professional development for teachers to determine how they might best use the data generate to improve instruction and meet the individual needs of their students. From the research on interim assessment programs, a constant and paramount goal is to improve instruction in the classroom. Ideally, as instruction improves student achievement will, likewise, be positively affected. Interim assessment data may target areas of need for student improvement. Using this information to individualize instruction to help all students should be the goal. Educators ranging from the classroom teacher all the way to the superintendent are bombarded with educational information all the time. There is no shortage of educational strategies and programs. Training teachers properly on the right programs should be of paramount importance to all school administrators. Simply selecting the right interim assessment program is only the first step. Training all members of the educational community on their roles of using the data may make the program successful for everyone involved.

Future Research Considerations

The findings of this study do not support that TLI or NWEA interim assessments provide an advantage over one another in mathematics achievement for sixth- and seventh-grade students. Additionally, the findings do not support that there exists a difference in mathematics performance based on gender. However, TLI assisted students

outperformed NWEA assisted students in literacy achievement for both sixth- and seventh-grade students. Furthermore, females outperformed males in literacy performance for sixth- and seventh-grade students according to the findings of this study. To evaluate the impact of specific interim assessments and gender on mathematics' and literacy achievement, the following studies are recommended:

1. A study examining the effects of interim assessments on mathematics and literacy achievement on different areas of Arkansas (Ex. Central or Eastern Arkansas)
2. A study examining the effects of the ACT-Aspire interim assessment program over the next several years due to the change to ACT-Aspire as the comprehensive state assessment
3. An examination of the impact of the use of interim assessment programs on classroom instruction
4. A study examining the effects of socio-economic status, ethnicity, and other potential factors on mathematics and literacy achievement
5. A study examining gender effects on mathematics and literacy achievement at different grade levels including elementary school, middle school, and high school students
6. An examination comparing multiple interim assessment programs for their effect on mathematics and literacy scores for multiple grade levels (elementary, middle, and high school)
7. A study on the effects of gender on academic achievement on different regions of the state and/or nation

The United States puts a high priority on public education. While the local and state governments have always had a prominent role in education, the federal government has injected itself in the arena in recent years with the initiatives including No Child Left Behind and Race to the Top. The Department of Education promotes student achievement to prepare young people for global competitiveness, fostering excellence in education and ensuring equal access. The use of interim assessments may be important in the goal to promote student learning.

The topic of interim assessment programs has become a common conversation for contemporary educators. Blazer (2008) stated that, while research is limited on the effect of interim assessments, there is an indication that using periodic assessments to guide instruction may have a positive impact on student achievement. For this study, the comparison of two popular interim assessment programs may provide new data that might be useful in choosing a program that might yield the best results for an individual district. Interim assessments are available at a wide variety of prices. Determining which program fits a districts' needs financially, student-wise, and subject-wise is important. In fact, a district might determine that purchasing an interim assessment program is not necessary at all for their district. However, school administrators must determine creative ways to improve student learning in their districts. Choosing the most suitable interim assessment program could be a step towards that goal.

The effect of gender on academic achievement has produced mixed results in research. Penner and Paret (2008) stated that, while much research indicates boys will outperform girls by middle school or the beginning of high school, some studies show that it occurs much earlier. For this study, there were no significant differences in

mathematics performance by gender for sixth- and seventh-grade students. Klinger et al. (2009) stated that female students' superiority in literacy areas has been relatively consistent in research for the past four decades. Consequently, this study indicated that females outperformed males at a significant level in literacy for sixth- and seventh-grade students in Northwest Arkansas. If certain challenges are a factor for specific genders regarding a specific academic subject, these issues need to be studied to determine how to overcome the barriers and succeed. The effects of gender on academic achievement must continue to be an important topic for educators. Challenges regarding gender can be met with the support of the entire school community.

REFERENCES

- Appalachia Educational Laboratory. (2005). *Research brief: Aligned curriculum and student achievement*. Retrieved from <http://www.edvantia.org>
- Arter, J. A. (2010). *Interim benchmark assessments: Are we getting our eggs in the right basket?* Denver, CO: NCME Conference.
- Axelson, M. (2005). *Online, standards-based, formative assessment conference proceedings* (Report). Retrieved from ERIC database. (ED489136)
- Babo, G., Tienken, C. H., & Gencarelli, M. A. (2014). Testing, socio-economic status, and the odds of passing Grade 8 state tests in New Jersey. *Research in Middle Level Education Online*, 38(3), 1-9.
- Blazer, C. (2008). *How interim assessment affect student achievement* (Report). Miami, FL: Miami-Dade County Public Schools.
- Brown, A. A., & Green, R. L. (2014). Practices used by nationally Blue Ribbon Award winning principals to improve student achievement in high-poverty schools. *National Forum of Applied Educational Research Journal*, 27, 2-15.
- Brown, R. S., & Coughlin, E. (2007). *The predictive validity of selected benchmark assessments used in the mid-Atlantic region*. Retrieved from ERIC database. (ED499099)
- Chappius, S., & Chappius, J. (2008). The best value in formative assessment. *Educational Leadership*, 65(4), 14-19. Retrieved from <http://www.asdc.org/>

- Coffey, H. (2010). *Benchmark assessments*. Retrieved from <http://www.learnnc.org/lp/pages/5317>
- Common Core State Standards Initiative. (2012). *Mission statement*. Retrieved from <http://www.corestandards.org/>
- Crooks, T. (2001, September). *The validity of formative assessments*. Paper presented at the British Educational Research Association Annual Conference, Leeds, United Kingdom.
- Demary, J. L., Magill, M. K., & Finley, C. W. (2000). *A study of effective practices in Virginia schools: Educator perspectives of practices leading to student success*. Retrieved from <http://clihome.com/Docs/CM/VirginaFullReport.pdf>
- Ebenuwa-Okoh, E. E. (2010). Influence of age, financial status, and gender on academic performance among undergraduates. *Journal of Psychology, 1*(2), 99-103.
- Eccles, J., Jacobs, J., & Harold, R. (1990). Gender role stereotypes, expectancy effects, and parents' socialization of gender differences. *Journal of Social Issues, 46*(2), 183-201.
- EHow. (2013). *Horizontal alignment definition?* Retrieved from <http://www.ehow.com/>
- Fleming, J. (2013). Interim assessments data: Guide your students, empower your practice. *The International Educator 28*(2), 17.
- Goertz, M., Olah, L. N., & Riggan, M. (2009). *From testing to teaching: The use of interim assessments in classroom instruction*. Retrieved from ERIC database. (ED519791)
- Guskey, T. R. (2003). Using data to improve student achievement. *Educational Leadership, 60*(5), 6-11.

- Hicks, M. (2014). *Four ways MAP or MPG interim assessments can support the transition to common core*. Retrieved from <https://www.nwea.org/blog/2014/four-ways-map-mpg-interim-assessments-can-support-transition-common-core/>
- Holmes, R. (2012, November 26). How does vertical alignment improve student performance in the schools? *The Holmes Education Post*. Retrieved from <http://theholmeseducationpost.com/>
- Hyde, J. S., & Mertz, J. E. (2009). Gender, culture, and mathematics performance. *Proceedings of the National Academy of Sciences*, 106(22), 8801-8807. doi:10.1073/pnas.0901265106
- Igbo, J. N., Onu, V. C., & Obiyo, N. O. (2015). Impact of gender stereotype on secondary school students' self-concept and academic achievement. *SAGE Open*, 5(1). doi:10.1177/2158244015573934
- Johnson, J. (2013). *Learning Point Associates: Data workshop*. Retrieved from <http://www.air.org>
- Kercheval, A., & Newbill, S. L. (2002). *A case study of key effective practices in Ohio's improved school districts*. Retrieved from <http://www.clihome.com/Docs/CM//OhioFullReport.pdf>
- Klinger, D. A., Shulha, L. A., & Wade-Woolley, L. (2009). *Towards an understanding of gender differences in literacy achievement*. Retrieved from http://www.eqao.com/en/research_data/Research_Reports/DMA-docs/literature-review-gender-differences.pdf

- Konstantopoulos, S., Miller, S. R., & Ploeg, A. V. (2013). The impact of Indiana's system of interim assessments on mathematics and reading achievement. *Educational Evaluation and Policy Analysis, 35*(4), 481-499.
- Kurz, A., Talapatra, D., & Roach, A. T. (2012). Meeting the curricular challenges of inclusive assessment: The role of alignment, opportunity to learn, and student achievement. *International Journal of Disability, 59*(1), 37-52.
- The Learning Institute (TLI). (2013a). *About TLI*. Retrieved from <http://www.tli.net/about-tli>
- The Learning Institute (TLI). (2013b). *Professional development*. Retrieved from <http://www.tli.net/what-we-do/professional-development>
- The Learning Institute (TLI). (2013c). *Reporting services*. Retrieved from <http://www.tli.net/what-we-do/reporting-services>
- The Learning Institute (TLI). (2013d). *TLI common core transition*. Retrieved from <http://www.tli.net/what-we-do/tli-common-core-transition>
- The Learning Institute (TLI). (2013e). *TLI literacy program*. Retrieved from <http://www.tli.net/what-we-do/tli-literacy-program>
- The Learning Institute (TLI). (2013f). *TLI mathematics program*. Retrieved from <http://www.tli.net/what-we-do/tli-mathematics-program>
- McGraw Hill Education. (2015). *Study confirms positive effects of acuity interim assessments for reading and mathematics*. Retrieved from <http://www.mheducation.com/about/blog/study-confirms-positive-effects-acuity-interim-assessments-reading-and-mathematics>
- No Child Left Behind Act of 2001. (2002). Pub. L. No. 107-110, § 115, Stat. 1425.

- Northwest Evaluation Association (NWEA). (2013a). *Educational assessment, student centered learning, common core assessments*. Retrieved from <http://www.nwea.org/>
- Northwest Evaluation Association (NWEA). (2013b). *Get to know NWEA*. Retrieved from <https://www.nwea.org/assessments/map/>
- Northwest Evaluation Association (NWEA). (2013c). *Home, assessments, standards alignment, Arkansas*. Retrieved from <https://www.nwea.org/assessments/standards-alignment/arkansas/>
- Northwest Evaluation Association (NWEA). (2013d). *Our history*. Retrieved from <http://www.nwea.org/node/165>
- Northwest Evaluation Association (NWEA). (2013e). *Our mission, vision, and values*. Retrieved from <http://www.nwea.org/node/130>
- Northwest Evaluation Association (NWEA). (2013f). *Practices and policies*. Retrieved from <http://www.nwea.org/node/1047>
- Ogutu, J., Peter, O., & Collins, M. (2011). Perceived effects of gender role stereotyping on academic aspirations of primary school pupils. *International Journal of Current Research*, 3(5), 135-142.
- Pallant, J. (2013). *SPSS Survival Manual* (5th ed.). New York, NY: McGraw-Hill.
- Penner, A., & Paret, M. (2008). *Gender differences in mathematics achievement: Exploring the early grades and the extremes*. Retrieved from www.sciencedirect.com

- Perie, M., Marion, S., & Gong, B. (2007). *The role of interim assessments in a comprehensive assessment system: A policy brief*. Retrieved from <http://www.ncica.org/publications/>
- Polikoff, M. S. (2012). Instructional alignment of No Child Left Behind. *American Journal of Education, 118*, 341-365.
- Questar Assessment. (2013). *2013 Arkansas Augmented Benchmark Examinations for Grades 3-8—Technical Report*. Apple Valley, MN: Questar Assessment.
- Reeves, D. B. (2006). *The learning leader: How to focus school improvement for better results*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Ryan, J. M. (2010). *Envisioning a state educational system: Improving learning through a comprehensive assessment system* (Report). Olympia, WA: Office of Superintendent of Public Instruction.
- Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research, 78*(1), 153-189. doi:10.3102/0034654307313795
- Southworth, R. A. (2008). *Formative assessment*. Retrieved from <http://www.edspeak.org/>
- Squires, D. (2012). Curriculum alignment research suggests that alignment can improve student achievement. *The Clearing House, 85*, 129-135.
- Taber, K. S., Riga, F., Brindley, S., Winterbottom, M., Finney, J., & Fisher, L. G. (2011). Formative conceptions of assessment: Trainee teachers' thinking about assessment issues in English secondary schools. *Teacher Development, 15*(2), 171-186.

- Vendlinski, T. P., Niemi, D., Wang, J., & Monempour, S. (2008). *Improving formative assessment practice with educational information technology*. Retrieved from <http://www.csa.ucla.edu>
- Vertical alignment*. (2013). Retrieved from <http://www.teacher2teacher.info>
- William, D. (2006, July). *Does assessment hinder learning?* Paper presented at the ETS Breakfast Salon, London, England.
- William, D. (2007). *What does research say the benefits of formative assessment are?* Retrieved from <http://www.nctm.org/>
- Zembar, M., & Blume, L. (2011). *Gender and academic achievement*. Retrieved from <http://www.education.com/reference/article/gender-academic-achievement/>
- Zimmerman, B. J., & Dibenedetto, M. K. (2008). Mastery learning and assessment: Implications for students and teachers in an era of high-stakes testing. *Psychology in the Schools*, 45(3), 206-216. doi:10.1002/pits.20291