Harding University
Scholar Works at Harding

Dissertations

5-2015

The Effect of Assessment Method on End of Course Geometry and Algebra Achievement

Miguel Hernandez
Harding University

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

Part of the Educational Assessment, Evaluation, and Research Commons, Educational Leadership Commons, and the Gender Equity in Education Commons

Recommended Citation

Hernandez, Miguel, "The Effect of Assessment Method on End of Course Geometry and Algebra Achievement" (2015). *Dissertations*. 36.
https://scholarworks.harding.edu/hu-etd/36

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.

THE EFFECT OF ASSESSMENT METHOD ON END OF COURSE GEOMETRY AND ALGEBRA ACHIEVEMENT

by

Miguel Hernandez

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

May 2015

THE EFFECT OF ASSESSMENT METHOD ON END OF COURSE GEOMETRY

AND ALGEBRA ACHIEVEMENT

by

Miguel Hernandez

Dissertation

Dissertation Advisor — Date 03/03/15

Dissertation Reader — Date 3/3/2015

Dissertation Reader — Date 3-3-15

Dean of the Cannon-Clary College of Education — Date 3/3/15

Assistant Provost for Graduate Programs — Date 3/3/15

© 2015

Miguel Hernandez

All Rights Reserved

ACKNOWLEDGMENTS

Many people contributed to the completion of this dissertation. My wife, Susan, provided constant encouragement and support throughout the entire process and gave me the strength I needed to finish. In addition to my wife, my dad and in-laws provided me with confidence that inspired me in times of need. My three daughters, Sophie, Olivia, and Josie, reminded me of the importance of the task, as I needed frequent reminders as to why I was going through this process.

The staff at Harding University has been outstanding. The guidance provided by Dr. Michael Brooks and Dr. David Bangs has been the foundation of my work, and without them, it would never have been possible. My Advisor, Dr. Usenime Akpanudo, has kept me on task to complete the assignments within the timelines and has coordinated with the rest of the Harding staff to ensure completion. My readers, Dr. Chaney Floyd and Dr. Clara Carroll, have been beneficial in their suggestions for improvement. I also want to thank my cohort members. They have always been there for me to lean on and inspire me to finish strong.

DEDICATION

I would like to dedicate this work to my family and friends. I also hope this work glorifies my Lord and Savior. My hope is that this will lead to aiding administrators in making the best decisions for children. The school business is more difficult than ever in this age of accountability. I thank all of my brothers and sisters in education for your hard work and dedication to our future.

ABSTRACT

by
Miguel Hernandez
Harding University
May 2015

Title: The Effect of Assessment Method on End of Course Geometry and Algebra Achievement (Under the direction of Dr. Usenime Akpanudo)

The purpose of this dissertation was to add to the existing research concerning the effects of assessment on mathematics achievement. The effects by gender or SES of students enrolled in school districts that used a commercial assessment versus school districts that used local assessments on mathematics achievement as measured by the end of course algebra I exam or end of course geometry exam.

This quantitative, causal comparative study was performed in six rural high schools in the Arkansas River Valley. The high schools had an approximate 700-student population of which 53% were categorized as free and/or reduced lunch and 51% were female. The end of course algebra I exam and geometry exam, given to all students enrolled in each course, was used as the instrument to measure mathematics achievement.

Included in the sample were all first time 9th graders for algebra I and first time 10th graders for geometry. Exactly 711 students comprised the sample. The students were classified according to their gender, SES, and the type of assessment method. The two categories of assessment were student enrolled in a course where The Learning Institute (TLI) interim assessment was used versus where a locally made assessment was used.

Four 2 x 2 factorial ANCOVA's were used to analyze the data for all hypothesis. No significant interaction effects were observed between students for assessment type and gender or assessment type and SES. For algebra achievement, there were significant difference found for assessment type but not for the main effects of gender or SES. For geometry, there were significant differences found for the main effects of assessment type, gender, and SES.

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER I

INTRODUCTION

School administrators struggle to find the next best idea that will lead to higher student achievement so that they can be openly held accountable in the public eye. Administrators are always bombarded with copious amounts of programs and tools that are advertised to deliver maximum results in order to curtail the consequences that are mandated by low student performance (Cobb & Rallis, 2008). Students are constantly being assessed, divided, taught, and reassessed. The school climate appears to vary with test scores. Questions linger in faculty minds whether there have been enough preparation on their part. Did the school use the right instructional methods and assessment strategies? Will the school be placed on a school improvement list for low benchmark achievement? There has been tremendous pressure placed on school administrators and teachers since the increase in accountability that came with the No Child Left Behind Act of 2001(NCLB, 2002). NCLB (2002) has led school administrators to seek out new and better ways to improve their student’s learning. The hope is that this will increase student achievement. According to Ananda (2003), current strategies mainly address curriculum alignment, integration of technology, and assessment. School administrators have also looked at many factors that may affect achievement on either mathematics or literacy. These factors include SES, gender, and how a student feels about mathematics (Myers, 1986).

According to Johnson, Arumi, Ott, & Remaley (2006), some school administrators have turned to interventions designed to increase the amount of class time spent on mathematics or science. One of the interventions are programs that have come into the market that promise to increase scores based on researched strategies. These are usually based on providing services such as curriculum alignment, interim assessments, research, consulting, and technology services to help teachers, administrators, and policy makers more effectively meet the needs of all students (TLI, 2010). Another possible intervention is required supplemental instructional services that are provided by school districts to increase achievement as mandated by lawmakers when a school has low test scores (Heinrich, Meyer, Whitten, & Urban, 2009). School administrators might also try to increase teacher quality through professional development. According to Georges, Borman, and Lee (2010), this intervention has been identified to be a critical part of student achievement. School administrators also spend a large amount of school funds on professional development each year to enhance the quality of teachers. Georges et al. reported,

> The differences between expectations for students and expectations for teachers mathematics content knowledge means that teachers may not have the content knowledge to teach to the standards required of elementary students. This gap means that students may not gain the mathematicsematical foundation necessary for later achievement in mathematics (p. 22).

This means that if teachers hold their students to high standards, then administrators must hold the teachers to high standards. The desire for increased student achievement has led many school administrators to interim assessment as a teaching tool that will aid the teacher in identifying student deficiencies and address academic needs (Black & Wiliam, 2009). Schools administrators have turned to the interim assessment programs that are being marketed by commercial assessment services that may or may not be effective. The questions that teachers often have are which assessments are better to help the students meet the goals of the No Child Left Behind Act? Dunn and Mulvenon (2009) stated,

> Many teachers do not feel empowered when dealing with assessment issues as there is a glaring absence of understanding in both the classroom and the literature with regard to how to fully use the power of both summative and formative assessments in education (p. 3).

One such commercial assessment service that claims to empower teachers to meet the challenges of assessment is TLI.

TLI has developed a system of curriculum planning, interim assessment, research, consulting, and support that is designed to help the teacher become better equipped to understand the level of learning going on in the classroom. TLI develops assessment products that are designed specifically to mirror the state assessments in both mathematics and literacy (TLI, 2010). The aim of this study is to examine the effectiveness of a commercially available assessment product (TLI) in improving student achievement in algebra I and geometry when compared to traditional assessment methods. In consideration of this, the following purpose statements have been developed for this study.

Statement of the Problem

First, the purpose of this study was to determine the effects of method of assessment by gender on algebra achievement for a sample of Arkansas students while controlling for eighth grade mathematics achievement. Second, the purpose of this study was to determine the effects of method of assessment by SES on algebra achievement for a sample of Arkansas students while controlling for eighth grade mathematics achievement. Third, the purpose of this study was to determine the effects of method of assessment by gender on geometry achievement for a sample of Arkansas students while controlling for algebra achievement. Fourth, the purpose of this study was to determine the effects of method of by SES on geometry achievement for a sample of Arkansas students while controlling for algebra achievement.

Background

School administrators work very hard to ensure that students have an environment where they can excel and perform at an optimal achievement level. The NCLB Act (2002) has given administrators the accountability of increasing student achievement. Because of this, many administrators have studied and adopted many instructional strategies in order to meet the requirements of NCLB. Curriculum alignment is a strategy that many administrators employ to raise student achievement in all populations and sub-populations identified by NCLB. Mitchell (1999) pointed out that curriculum alignment had positive effects on student achievement in mathematics. Aligning the curriculum is a strategy that school administrators can adopt in their school setting that can have an immediate impact and limit the effects of poverty, race, gender, and school size.

Lesisko, Wright, and O'Hern (2010) stated that technology integration in classroom instruction is something that school administrators are employing to enhance student learning and increase achievement. School administrators must become the technology leader in order to advance their school to higher achievement levels.

Technology by itself will not increase achievement. There must be a program of professional development for teachers to know how to use the technology in order to increase achievement. According to Cobb and Rallis (2008), district personnel with local control and the ability to make its own choices feel empowered to help students achieve. In contrast, schools that have high external regulation may have staff that may have low morale. School leadership has been targeted by lawmakers as a large part of what makes a school successful. Sawchuk (2008) noted that there is a need for stronger focus by district leaders to be the leader of school reform. School leaders must look at the wide array of factors besides curriculum and technology that can affect student achievement. SES is a research-based factor that can contribute to how a student achieves in mathematics and literacy. Myers (1986) stated that students that are in a high poverty school are less likely to have high achievement compared to students in low poverty schools. Some studies have shown that there is a gap in the problem solving scores between male and females. Wilson and Zhang (1998) reported that,

> The results in some ways contradict the more hopeful conclusions of other studies that have shown that the gender gap is narrowing, though they do affirm some of the results of the Hyde et al. study that showed males stronger in problem solving at the high school years. The results suggest that, while the gap may be narrowing on traditional multiple-choice tests, it is still present on more complex items that require students to construct their own responses and communicate their thinking. It is especially disturbing to see that gap increases with grade level, which is in keeping with earlier studies showing females falling behind in adolescence. (p. 12)

Therefore, gender and how student perception about achievement are also factors that affect student performance.

Brown and Hirschfield (2007) pointed out that student perception should be taken into account when studying the effectiveness of any learning tool. According to TLI (2010), students perceive their method of assessment and feedback as being beneficial to their learning process. “The evidence here is that mathematics scores increase if students agree that assessment itself makes students accountable for learning” (Brown & Hirschfield, p. 71). TLI offers the tools that give the teacher a simple and immediate way to show what levels the student or class is achieving.

Schools use many other targeted strategies to combat the factors that lead to low achievement in mathematics. These are increasing time spent on mathematics, supplemental services, increasing teacher quality, and assessment techniques. According to Johnson et al. (2006), business leaders claim that students are not sufficiently skilled and knowledgeable about mathematics and may need more instruction than what is currently given. This has lead school administrators to increase the amount of mathematics courses that students are now required to take for graduation. The effects of this increase remain to be seen due to it being newly implemented. Teacher quality is another major factor that is being scrutinized by school leadership. “Since teacher quality has emerged as one of the most powerful variables in student success, the focus of policy reform must be on building the capacity of our teachers to meet the challenges our schools face” (National Board Resource Center at Stanford University, 2010, p. 1). School leaders must do a great job of finding and training quality teachers to improve the instruction going on in the classroom. Quality teachers are effective in using assessment to gauge the student learning. Interim assessment is a tool that teachers often use as a part of their comprehensive educational program.

Although the existing literature points to the deep impact that interim assessment can have on the effectiveness of student achievement, a scarcity of research exists on any of the particular companies that offer a system for interim assessment. For this reason, educational administration communities are unaware of the best way to assist teachers' use of assessment to better student achievement. To address this gap in the literature, this particular study was conducted to specifically explore the outcomes of student achievement between using a commercial system of interim assessment as compared to traditional assessment methods.

There are often differences of opinion on what an interim assessment is and is not. Interim assessments must be something that is practiced regularly in the classroom. Cech (2008) noted that schools have spent in the billions of dollars on assessments that cannot be considered interim. "That's a lot of money being spent on something that experts say can't really be sold and only practiced" (p. 2). According to Black and Wiliam (2010), interim assessment is a conceptual idea that is often used in a wide array of instructional systems.

High test scores are beginning to be a leading attraction for parents that are looking for schools to put their children in. There is starting to be a competition between schools for students and that translates into dollars for the school. This is visible by the increasing amount of advertisements that are presented by school officials in newspapers, television commercials, and billboards. "NCLB does indeed create a context for examining assessment because in many settings the pressure to raise test scores has created overnight celebrity status for assessment" (Shepard, 2009, p. 36). According to Ginsburg (2009), interim assessments are best described in three parts: observation, testing, and clinical interview. The idea of this paper is to look at the testing portion. Ginsburg also stated that one could learn many things from a test. They can be used to obtain information about how much a student understands the material. The test can also show the teacher what common misunderstandings each or all the students have about a concept. This can then be used to refine a teaching strategy. According to Dunn and Mulvenon (2009), interim assessments have many positives that will directly influence teachers and students.

Perie, Marion, and Gong (2009) stated that school leaders should be aware of the different types of assessments that are available. They must also understand that companies that are hurriedly flooding the market place with assessment systems have been far removed from the everyday business of the classroom teacher. These companies often promise big gains and superior systems, but usually come up short. "A good interim assessment can be an integral part of a state's or district's comprehensive assessment system, used in conjunction with classroom formative assessments and summative end-of-year assessments" (p. 13). However, teachers often struggle with having time to use test data to help them plan curriculum for the current class. The data gained from end of year assessments only relates to the students that are leaving this class for a new one. It is difficult to change curriculum for the next year based on last the previous year students. This is also very true in the mathematics classroom. Davis and McGowen (2007) noted that mathematics teachers would not have time to take testing data and use it to alter the curriculum immediately to help the currents students to meet their specific needs. TLI tries to aid the teacher in this aspect by offering reporting services for multiple-choice questions that can be returned to the teacher instantly. The teacher can then use the organized reports to identify misconception by individual students or problems with the curriculum itself.

According to Ayala et al. (2008), to embed assessments into the existing curriculum, the following needs to happen: collaboration between assessment specialist and curriculum developers, adequate professional development for teachers using the program, embedded assessments need to be linked to the overall goal of the curriculum, student understanding must be tracked and feedback must be given to students, assessment pedagogies must be understood, and frequency of assessments must be considered. TLI uses their assessment specialists to work with the district curriculum teams to help develop pacing guides and assessment frequency. TLI also offers professional development for teacher. This professional development consists of making the teacher comfortable with the assessment structure, how to analyze reports, and use their online tools.

School administrators have seen the research that details many of the indicators that prevent students from obtaining optimal achievement. Administrators have also viewed the research on some of the interventions that can be used to improve student achievement. One of these interventions that will be studied further will deal with assessment. Students and teachers often view many classroom assessments as impeding the process of learning. TLI offers through its mathematics program many aspects that allow the student to be informed by using immediate feedback that shows the student and teacher strengths and weaknesses. This can then be used to re-teach using a variety of methods. These immediate feedback sessions have been linked to creating a positive perception of assessment by students. This positive perception may lead to a better learning environment and increase student achievement.

Hypotheses

The researcher generated the following null hypotheses:

$H1_a$: There will be no significant interaction between type of assessment and gender on algebra achievement when controlling for eighth grade mathematics achievement.

$H1_b$: There will be no significant difference in the main effect of type of assessment on algebra achievement when controlling for eighth grade mathematics achievement.

$H1_c$: There will be no significant difference in the main effect of gender on algebra achievement when controlling for eighth grade mathematics achievement.

$H2_a$: There will be no significant interaction between type of assessment and SES on algebra achievement when controlling for eighth grade mathematics achievement.

$H2_b$: There will be no significant difference in the main effect of type of assessment on algebra achievement when controlling for eighth grade mathematics achievement.

$H2_c$: There will be no significant difference in the main effect of SES on algebra achievement when controlling for eighth grade mathematics achievement.

$H3_a$: There will be no significant interaction between type of assessment and gender on geometry achievement when controlling for algebra achievement.

$H3_b$: There will be no significant difference in the main effect of type of assessment on geometry achievement when controlling for algebra achievement.

$H3_c$: There will be no significant difference in the main effect of gender on geometry achievement when controlling for algebra achievement.

$H4_a$: There will be no significant interaction between type of assessment and SES on geometry achievement when controlling for algebra achievement.

$H4_b$: There will be no significant difference in the main effect of type of assessment on geometry achievement when controlling for algebra achievement.

$H4_c$: There will be no significant difference in the main effect of SES on geometry achievement when controlling for algebra achievement.

Description of Terms

Eighth grade mathematics achievement. This is a measure of students' mathematics ability. For the purposes of this study, it was defined as the scale scores that are achieved by students on the Arkansas eighth grade benchmark mathematics exam. The 2008-2009 scores were used for eighth grade mathematics achievement.

Algebra achievement. This is a measure of students' mathematics ability. For the purpose of this study, it was defined as the scale scores that are achieved by students on the Arkansas algebra I end of course exam. The 2008-2009 and 2009-2010 scores were used for algebra achievement.

Commercial assessment programs. This is any company that develops and produces assessment and instructional supplements for a profit. Examples include TLI and Northwest Evaluation Association. For the purposes of this study, TLI was used as this type of assessment.

Geometry achievement. This is a measure of students' mathematics ability. For the purposes of this study, it was defined as the scale scores that are achieved by students on the Arkansas geometry end of course exam. The 2009-2010 scores were used for geometry achievement.

Interim assessment. Perie et al. (2009) defined interim assessment as medium-scale, medium cycle assessments, falling between summative and formative assessments and are usually administered at the school or district level. For the purposes of this study, these were the assessments used by the TLI assessment program.

SocioSES (SES). SES is an economic and sociological combined total measure of a person's work experience and of an individual's or family's economic and social position relative to others, based on income, education, and occupation. For the purposes of this study, SES was defined as a student's status with regards to the Federal Free/Reduced lunch program. Students who are participants in the program were considered to be of Economic Disadvantaged, and students who were non participants were considered to be Non-Economic Disadvantaged.

The Learning Institute (TLI). TLI is a for profit company that provides schools with curriculum alignment and support, interim assessments, research, consulting, and technology services to help teachers, administrators, and policymakers more effectively meet the needs of all students. For this study, the term TLI was used to represent the interim assessment method provided by this company.

Traditional methods of assessment. This was defined as all other current assessment strategies, including interim, formative, and summative assessments that are produced by teachers at the local education agency. Specifically in this study, the traditional method of assessment was used in reference to assessment methods developed by local school districts. Daily, chapter, quarterly, and other locally developed assessments are examples of this type of assessment.

Significance

Making decisions about the distribution of funds can be a very challenging task for school administrators. There are also decisions to be made about the instructional tools that are purchased for the teachers to help instruct students. School leaders are reluctant to invest in programs or learning strategies that are not effective for the majority of students. This study is therefore significant because it examines TLI's approach to helping teachers by providing tools such as curriculum maps, quiz builders, and formative assessments with data analysis. The findings of this study will help school leaders to make informed decisions about whether or not TLI program is effective for algebra and geometry high stakes testing. The study is also significant because it will also look into the relationship between mathematics achievement and gender. This study will look at the relationship between mathematics achievement and SES.

Process to Accomplish

Design

A quantitative, causal comparative strategy was used for this study. The independent variables for the first set of hypotheses were method of assessment and gender. The dependent variable was algebra achievement. For these hypotheses, eighth grade mathematics achievement was used as a covariate. The independent variables for the second set of hypotheses were method of assessment and gender. The dependent variable was geometry achievement. For these hypotheses, algebra achievement was used as a covariate.

The independent variables for the third set of hypotheses were method of assessment and SES. The dependent variable was algebra achievement. For these hypotheses, eighth grade mathematics achievement was used as a covariate. The independent variables for the fourth set of hypotheses were method of assessment and SES. The dependent variable was geometry achievement. For these hypotheses, algebra achievement was used as a covariate.

Sample

The sample for this study consisted of 711 students who took the end of course geometry and algebra exams in six Arkansas schools in the River Valley area of the state. Three schools where chosen based on their initial participation in using TLI during the 2009-2010 school year. Although these schools had used other traditional methods of assessment that included those developed by their local district during the 2008-2009 school year, they had switched to TLI during the 2009-2010 school year. A comparison sample was also drawn from three different schools that used traditional methods of assessment developed by their local district during the 2008-2009 and 2009-2010 school years. All the schools from which samples were drawn shared similar demographic characteristics. These demographics included student to teacher ratio, percentage of free or reduced lunch, ethnic makeup, and gender percentages. All students sampled for this study were traditional 9th graders when they took algebra I and were also traditional 10th graders when they took geometry. The students were chosen based on both random and convenience techniques from the total population in each school that met the above criteria.

Instrumentation

The Arkansas Comprehensive Testing, Assessment, and Accountability Program (ACTAAP) end of course examination will be the instrument for this study. According to the Arkansas Department of Education (2010), ACTAAP includes an End-of-Course Examination in geometry and algebra I. It consists of multiple-choice and open-response questions that directly assess student knowledge. The Arkansas Geometry and Algebra I Mathematics Curriculum Framework is the basis for development of the End-of-Course Examination. The geometry exam consists of eight sessions over a 2-day period. The first session consists of 20 multiple-choice questions. The second session consists of 15 multiple-choice questions. The third session consists of three open response questions. The fourth session consists of 20 multiple-choice questions. The fifth session consists of 15 multiple-choice questions. The sixth session consists of two open response questions. The seventh session consists of 20 multiple-choice questions. The final session consists of two open response questions.

The algebra I exam consists of eight sessions over a 2-day period (Arkansas Department of Education, 2010). The first session consists of 20 multiple-choice questions. The second session consists of 15 multiple-choice questions. The third session consists of three open response questions. The fourth session consists of 20 multiple-choice questions. The fifth session consists of 15 multiple-choice questions. The sixth session consists of two open response questions. The seventh session consists of 20 multiple-choice questions. The final session consists of two open response questions.

Data Analysis

To address the first group of hypotheses ($H_{1a, b, c}$), a 2 x 2 factorial analysis of covariance was conducted to examine the interactions and main effects of method of assessment and gender as the independent variables and the algebra achievement as the dependent variable and eighth grade achievement as the covariate. To address the next three hypotheses ($H_{2a, b, c}$) , a 2 x 2 factorial analysis of covariance was conducted examine the interactions and main effects of type of assessment and gender as the independent variables and the geometry achievement as the dependent variable and algebra achievement as the covariate. To address the third set of hypotheses($H_{3a, b, c}$), a 2 x 2 factorial analysis of covariance was conducted examine the interactions and main effects of type of assessment and SES as the independent variables and the algebra achievement as the dependent variable and eighth grade achievement as the covariate. Finally, to address the last three hypotheses ($H_{4a, b, c}$), a 2 x 2 factorial analysis of covariance was conducted examine the interactions and main effects of type of assessment and SES as the independent variables and the geometry achievement as the dependent variable and algebra achievement as the covariate A two-tailed test with a .05 level of significance was used for all the tests in this study.

CHAPTER II

REVIEW OF RELATED LITERATURE

Educators are constantly looking for ways to improve student achievement. Over the years, vendors have developed a variety of commercial products aimed to facilitate and enhance classroom instruction. Davis and McGowen (2009) stated that there are many limitations to what a commercial program can do for a teacher in the area of student achievement. Assessment programs are important in the area of education. Dunn and Mulvenon (2009) pointed out there is ample research that supports the use of assessment to improve student achievement. Mathis (2004) stated the NCLB Act of 2001 has raised the accountability measure imposed on school districts by putting a larger emphasis on state mandated assessment and interim assessment. To help teachers meet the accountability of NCLB and increase student achievement, many school district leaders have used TLI (2010) along with other interim assessment models.

Pressure to Improve Instruction

A study by Mathis (2004) revealed that NCLB is a federal government plan for education. The plan does not align with what it was intended to do. The Arkansas Department of Education (2010) stated that students are assessed for proficient learning and the scores are used to label a school as good or bad. Mathis noted that numerous fallacies exist with the standards-based Adequate Yearly Progress measures of the NCLB Act. The legislation puts significant pressure on school administrators to ensure that

100% of students must be proficient by a certain deadline (Keegan, Orr, & Jones, 2002). "Obviously, even the highest-performing schools may eventually find they cannot ensure that every student will reach a high standard" (Mathis, 2004, p. 144). The schools have the added pressure of having the sole responsibility of ensuring that the students have everything they need to be successful according to Cobb and Rallis (2008). Mathis (2004) noted that there is no aid for the school in making sure that the students have support and motivating factors from home.

Keegan et al. (2002) stated that each student does not have the same starting place when they enter the doors of the school and each grade level every year, and it is a struggle to give each child the individual time they may need to become proficient. Mathis (2004) noted the following:

> It is easy to predict that NCLB will fail for the very simple reason that it cannot succeed, but in politics strange transformations take place. Even the political forces that aligned to create NCLB are growing aware of the law's shortcomings and the unpopularity of many of the act's provisions. Even though hard-liners say "no amendments," we can reasonably expect to see the law repealed or transformed amid considerable political tacking and spinning. The important question is not how NCLB will or will not be brought into conformity with reality, but how we should transform American education in the aftermath. (p. 150)

Cobb and Rallis (2008) pointed out that schools are able to better meet requirements when they come from internal expectations rather than external expectations. "We propose that districts that operate under a balance of internal, external, and lateral accountability are more capable of changing, more capable of aligning purposes, values, and actions, and inherently more democratic institutions" (p. 199). Mathis (2004) noted the self-motivated student could excel with minimal directions from the teacher, as should schools.

A study by Sawchuck (2008) showed that there is a big disconnect between how district leadership and teachers view NCLB. There is a higher emphasis on accountability in schools and pressure to raise student achievement according to Keegan et al. (2002). The data collected by Sawchuk (2008) showed that there is limited influence by administrators on teacher practices in the classroom. Some of this can be attributed to the school administrator not having the time or knowledge to help each individual teacher align their specific standards to the state assessments. Sawchuk found the following:

> Increasingly, principals are expected to serve as instructional leaders in their schools, helping translate standards-based changes into practice and coalesce teachers into high-functioning teams. Not all principals have benefited from the preparation and professional development needed to play that role effectively. (p. 3)

Mathis (2004) noted that administrator quality might lead to them not being able to help a teacher in a specific area due to a lack of knowledge in a certain subject area.

Dunn and Mulvenon (2009) stated that there are both positive and negative effects of the NCLB Act. A negative effect is that more teachers are reporting increased levels of angst in dealing with the assessments and consequences of low performance. Seed (2008) pointed out that teachers are struggling with the pressure to improve student achievement and need the help of all stakeholders to be successful. Dunn and Mulvenon (2009) found the following:

> In the wake of NCLB, most teachers are experiencing high anxiety in the high-stakes testing era. However, the purpose of NCLB was not to inject fear into teachers; the purpose was instead to inject data-driven decision making into schools. A possible cause for the fear and anxiety teachers experience with regard to high-stakes mandated summative exams is rooted in the failure of assessment-related language and research to provide an effective model for improving teaching and learning through the use of state-mandated assessment data. (p. 3)

According to TLI (2010), one of the ways to alleviate some of the fears of NCLB is the implementation of an interim assessment program. This will allow the teacher to monitor the achievement throughout the year rather than waiting until the end of the year. The interim assessment will help the teacher make data-driven decisions about the instruction of his or her classroom as noted by Black and Wiliam (2010).

Sunderman, Orfield, and Kim (2006) stated that NCLB has changed the relationship between the federal government, state government, and local schools in relation to educating students. There has been more pressure put on school leaders to ensure that their schools measure up to the external accountability according to Mathis (2004). The belief is that this external accountability and sanctions will force teachers to improve instructional practices and raise student achievement. (Cobb & Rallis, 2008). Sunderman, Orfield, and Kim (2006) noted the following:

> By relying on the threat of sanctions and market mechanisms— choice and supplemental educational services—to force school improvement, the law tends to

> place the principals of low-achieving schools in the role of trying to produce very large gains every year for every subgroup of students. (p. 20)

Seed (2008) pointed out that NCLB was very prescriptive in its approach to increasing student achievement, but it neglected taking a collaborative approach with educators. One of the biggest assumptions made by NCLB, noted by Sunderman et al. (2006), is that teachers will begin to produce positive results if they feel more pressure by receiving sanctions and labeled as failures. The goal should be to attract and retain good teachers (Seed, 2008). These good teachers do not often want to go to failing schools where they have lower achieving students, limited resources, and limited support according to Sunderman et al. (2006).

Methods and Obstacles to Increase Mathematics Achievement

Committees that are outside of the school setting set the mathematics achievement standards, which are too high according to Mathis (2004). Curriculum alignment is a method of matching these mathematics standards to the teaching and assessment as noted by Seed (2008). “Alignment usually means only that the tests are not grossly incompatible with the standards. It does not mean that they comprehensively, validly, and reliably measure the performance for which the standards were set” (Mathis, 2004, p. 145). Curriculum alignment is set up by states to ensure that all students will eventually reach the state defined proficient level (Seed, 2008). School administrators must make sure that there is no misalignment between teacher or grade levels. This could cause severe gaps in a students’ knowledge (Ananda, 2003). “Alignment can be achieved through the use of sound standards and assessment development practices that focus on alignment during each step of the process” (Ananda, 2003, p. 6). Curriculum alignment is the first step in setting conditions that will improve teaching which should result in higher student achievement as according to Seed (2008).

Mitchell (1999) stated there are three steps to the curriculum alignment process. The first step identifies what the school district wants to accomplish. The second step requires a coordination with the textbook and the assessment instrument. The third step requires the creation of meaningful materials throughout the grade levels to fill in the gaps created by any misalignment of the textbook and instrument used for assessment. Strong curriculum alignment will make the assessment more valuable to the teachers as noted by Ananda (2003). Osta (2007) stated that culture is the biggest problem to designing a common assessment that is based on curriculum alignment which is useful to everyone. Seed (2008) noted that teachers should be provided with enough time to collaborate and form assessments that have minimal inequality for test takers. Osta (2007) found the following “it is virtually impossible to build a fair exam that would be equal among all cultures. For example, would it be fair to question a student living in a large metropolitan area about the volume of a grain silo without explaining the shape of such object” (p. 193)? Technology in the classroom has shown great advantages to helping students understand mathematics concepts. (Helen, 2011) A study conducted by Lesisko et al. (2010) said that school districts should be able to produce a group of students that are savvy in the use of technology. According to Helen (2011), technology can be used to enhance student learning and by correlation increase test scores. School administrators are very aware of the need for strong leadership in the area of technology and its aid in instruction and assessment as noted by Lesisko et al. (2010).

Helen (2011) pointed out that technology incorporated into the classroom could have many positives on mathematics achievement. An instructor will be more capable of understanding how to alter the pace of instruction quickly when using technology with assessment according to TLI (2010). Helen (2011) noted, “Many different hardware, software, and web-based tools can offer new approaches for teaching and learning mathematics. However, having the technologies available does not mean that teachers and students understand how to use them effectively, or even choose to use them at all” (p. 60). The students will be more engaged due to the interactive setting of the classroom (Strasser, 2010).

The final benefit given by Strasser (2010) is that the instructor will be able to provide the student with feedback in a less time before the use of technology. According to Helen (2011), instructors can use this information to direct instruction and students will be able to focus on their deficiencies. Strasser (2010) noted the negative aspects of the technology are often reliability of the technology, access issues, and time it takes to set up initially. Many teachers do not wish to spend the start-up time to get a new technology online according to Helen (2011).

Bellamy and Matvio (2010) noted that the classroom is the best place to reinforce understanding of material with the use of technology. Strasser (2010) noted that teachers should have a classroom setting that is able to guide the students in the use of technology for conceptual learning of science and mathematics. “Real-life demonstrations and experiences with mathematics and science principles are best learned in a technology education setting. Students will both learn and retain the concepts through applying them in a practical setting" (Bellamy & Matvio, 2010, p. 28). According to Gasbarra and Johnson (2008), poverty is a barrier to access to technology for some students.

Myers (1986) found that schools with high concentration of poverty are more likely to have black or Hispanic students that speak languages other than English and have low achievement. Students that are from poverty stricken schools tend to have lower achievement in the area of science and mathematics as noted by Gasbarra and Johnson (2008). Students in high concentration of poverty generally have lower achievement than this in low concentration of poverty after taking student and family characteristics, which can sometimes lead to an environment that is less than ideal and can lead to lower achievement (Myers, 1986). These differences are usually seen in literacy, but mathematics variability is most evident in the early grades as noted by Gasbarra and Johnson (2008). Gender, in addition to poverty is another indicator of mathematics achievement according to Scherer (2010).

According to Wilson and Zhang (1998), research contradicts other studies that show the achievement gap between male and females is narrowing. According to Phillips and Meloy (2012), males appear to be stronger in problem solving in the high school years. The gap between genders in multiple choice seems to be narrowing, but constructed response the gap is still present as noted by Wilson and Zhang (1998). Phillips and Meloy (2012) stated that the mathematics gender gaps are not as prevalent in the pre-k grade, but become more pronounced in the middle and high school grades. The gap also shows females falling behind in adolescence (Wilson & Zhang, 1998). In the middle grades, student perception of teacher caring has a greater effect on mathematics achievement than gender differences according to Strobel and Borsato (2012).

Brown and Hirschfeld (2007) noted that student perception of assessment plays a large role in their achievement level. The results showed that students that take part in their learning and proactively use feedback achieve more as noted by Strobel and Borsato (2012). Mathematics achievement is higher if the student believes that the assessment itself makes the students accountable for learning (Wilson & Zhang, 1998). Brown and Hirschfeld's (2007) research also showed that ethnicity, SES, and gender played a role in the achievement levels of mathematics. A study done by Johnson et al. (2006) showed that students want schools to prepare them for the workforce, but feel that they would be unhappy in careers relating to mathematics or science. Wilson and Zhang (1998) pointed out the gender differences in how student achieve in mathematics is something that is widely publicized, but students have virtually no different feeling about mathematics and science. "When students are asked about a variety of possible problems at their schools, concerns about the lack of emphasis on science and mathematics is near the bottom of the list" (Johnson et al., 2006, p. 8). Rutherford et al. (2010) noted that school leaders are looking at every avenue to increase mathematics achievement. Strasser (2010) pointed out that technology is an equalizer in closing the achievement gap in mathematics. A study by Rutherford et al. (2010) showed utilization of interactive mathematics software that provides an individualized delivery of standards-based mathematics by using the fact that most basic mathematics concepts can be understood pictorially. Mathematics concepts are easily understandable to students when being taught using interactive technology according to Strasser (2010). The study showed that there is an increase in student achievement as it relates to gender differences and the gap in SES by using technology (Rutherford et al., 2010). Technology along with supplemental education services are another avenue for increasing mathematics achievement as noted by Strasser (2010).

According to Henrich, Meyer, and Whitten (2009), school leaders and parents are employing private providers of supplemental education services to increase achievement in mathematics. White, Loker, March, and Sockslager (2009) stated that in supplemental services in some cases to be effective for students that are interested in higher achievement, but the students are not required to sign up. The schools are only required to provide. These supplemental education service companies are privately owned and do not have a lot of oversight in the quality of instruction provided to students according to Heinrich et al. (2009). Students that are from low SES tend to fill up these seats. The students with high absentee rates are usually not going to attend (White et al., 2009). According to Heinrich et al. (2009), supplemental services has shown a low impact outcome for mathematics achievement due to lack of oversight by teachers, information to parents about the programs, and lack of interest by students.

A study by National Board Resource Center at Stanford University (2010) showed that school leaders are also looking at the quality of teachers as a measure of mathematics achievement. According to Lee, Robinson, and Sebastian (2012), teacher quality has a direct impact on a student's mathematics achievement. Georges et al. (2010) noted that elementary teachers are not getting sufficient content knowledge in the area of mathematics while is college to be effective in increasing student achievement. There is a lack of rigor in what it is expected of the students by the teacher as noted by Lee et al. (2012). Georges et al. (2010) claimed that the average classroom teacher does not have the depth of content knowledge to adequately teach rigorously. The quality and organization of professional development has also found to be lacking as a means to increase teacher understanding of mathematics content knowledge as noted by National Board Resource Center at Stanford University (2010). Georges et al. (2010) also pointed out that poor teacher preparation could lead to students not getting the foundation in elementary school that is necessary for later achievement. Some states are calling for more content knowledge intense policies for mathematics teachers according to National Board Resource Center at Stanford University (2010). It is not practical for elementary teachers to major in mathematics, but it is possible to use professional development to better their understanding of the mathematics concepts. (Georges et al., 2010). Teacher quality can become less of a factor in student achievement if there is a comprehensive assessment system used as noted by Lee et al. (2012).

Interim Assessment

According to Black and Wiliam (2010), one of the areas that teachers can improve student achievement is using effective assessment and feedback. Chappuis, Chappuis, and Stiggins (2009) pointed out that students need effective assessment in helping to analyze their problem areas. Interim assessment can be way for students to assess their own learning and gain feedback from the teacher on how to improve their mathematics skills Black and Wiliam (2010). TLI (2010) stated that it is more beneficial to students to give shorter and more focused tests at frequent intervals, than to give longer tests that are not in a timely manner. Black and Wiliam (2010) noted that there must be quality question items given within one week of learning new material. The increase in mathematics achievement by raising standards through assessment can only come about by getting buy-in from the teachers and students according to Chappuis, Chappuis, and

Stiggins (2009). Interim assessment is an important part of raising achievement levels in the classroom (Black & Wiliam, 2010).

Cech (2008) pointed out that interim assessments can be best defined by what they are not. They are not the long, year-end, state-administered, standardized, NCLB Act-required exams that are referred to as summative assessments. Black and Wiliam (2010) stated that they are also not assessments that are given in the middle of the year. Interim assessment is when the evidence is actually used to adapt the teaching to meet student needs (Cech, 2008). Interim assessments are more than a half-billion dollar business in the United States in the 2006-2007 academic year according to Black and Wiliam (2010). Chappuis et al. (2009) felt that testing companies are being too liberal with the label of interim assessment. This adds to the confusion of what a true interim assessment is and is not as noted by Cech (2008).

According to Sheppard (2009), the NCLBAct has increased accountability on teachers and school administrators. Cech (2008) noted that this pressure has created an overnight popularity for interim assessment. Chappuis et al. (2009) stated that interim assessment makes promises to increase student achievement levels, without the research to support these claims. Interim assessment is of little use to teachers if they do not know what to do if students cannot grasp a concept according to TLI (2010). Shepard (2009) noted that interim assessments are different from day to day formative assessments. Interim assessments must not be taken for granted because something is needed to diagnose individual students during the current year. Cech (2008) states that interim assessment is something that yearlong summative assessments cannot offer.

A study by Ayala et al. (2008) showed there are several considerations that should be taken into account in creating interim assesments. Chappuis et al. (2009) stated that collaboration between the assessment specialists and curriculum staff is important in creating a quality assessment. Interim assessments need to appear seamless to students and teachers according to Ayala et al. (2008). Cech (2008) noted that professional developments should be available to teachers so that they have a thorough understanding of interim assessment and how to use it in the class setting. According to Ayala et al. (2008), assements need to be linked to the overall goal of the curriculum and not just the material covered for that specific period of time. Learning trajectory should be used to help teachers track student understanding and provide feedback (Cech, 2008). The frequenecy and quantity of interim assesment should be set based on the needs of achieving the goals of the curriculum as noted by Ayala et al. (2008).

Commercial Interim Assessment Programs

Perie et al. (2009) stated that many commercially sold interim assessments claim that research show powerful gains by using interim assessment for student learining. It is not in the best interest of the school district to spend limted resources on assesments that are may not lead to increased student achievment as noted by Ayala et al. (2008). The assessments should be based on instructional goals and used to give teachers useful information according to Perie et al. (2009). A school that spends money on an interim assesment system should provide experiences that are not available of any state given assessment or local assessment Cech (2008).

According to Davis and McGowen (2007), teachers may not have to time in class to sift through the student data in a timely manner to change the curriculum to meet the needs of students. Technology combined with assessment can be an invaluable tool in giving the student timely feedback as noted by Helen (2011). This time lapse will make curriculum changes have to happen later, thus slowing down the effectiveness of the assessment shown by Davis and McGowen (2007). Perie et al. (2009) stated that, when the teacher does have time to analyze data, the teacher could become more aware of the changes that need to be made. “This can be done by more informed questions choices and instructional design, which can help transform their thinking on instruction that is better aligned to student needs” (Davis & McGowen, 2007, p. 28).

TLI (2010) stated that they provide interim assessments in mathematics, literacy and science. Their assessments help to determine strengths and weaknesses in curriculum and student knowledge. TLI provides immediate feedback through a variety of reporting services online so teachers and administrators can find the immediate needs of their students. TLI staff also works closely with member districts throughout the process by providing professional development, curriculum support, and intervention strategies to benefit all students.

According to TLI (2010), they work closely with districts to develop a clear and concise map that links learning expectations taught in the classroom to what is asked on the assessments. The curriculum maps are unique to each district. TLI provides guidance and support, but each district representative makes all decisions. Each district can decide how many assessments are to be given. Alignment is reviewed annually and can be changed year to year. TLI noted that they develop mathematics assessments for grades 1-8, algebra I, geometry, and algebra II. Each assessment includes up to 20 multiple-choice questions and one open-response question. There can be up to eight assessments per year.

TLI curriculum specialists approve all items for content and standard correlation. TLI replaces the mathematics item back each year, and does not use state released items. The old items are put into a quiz builder tool that allows teachers to use them for remediation and skill building.

TLI (2010) claims they are able to provide a single point of access to all district achievement data so that teachers and district leaders can have the right information. Reporting services are provided on-line where tests, answer sheets, and reports can be accessed. Benchmark data is provided with a breakdown of student performance in a school or district by subpopulation. Each module assessment report gives item-by-item analysis that allows teachers to explore areas of strength and weakness by student, class, or district. Teachers can give specific remediation in a timely fashion.

Conclusion

Since the implementation of NCLB legislation, there has been added pressure on teachers, school administrators, and students to increase mathematics achievement (Mathis, 2004). School leaders have studied the key components of NCLB in order to understand the accountability placed on mathematics achievement scores (Keegan et al., 2002). School leadership has tried a myriad of educational practices to help increase student achievement (Davis & McGowen, 2009). These methods have included curriculum alignment, technology, improving the quality of teachers, and assessment methods (Georges et al., 2010; Seed, 2008; Strasser, 2010; Wiliam & Black, 2009).

Curriculum alignment has been a very effective tool in increasing student achievement in mathematics due to teachers being able to match standards to assessment (Keegan et al., 2002). Technology in the classroom has shown increase student achievement by allowing students to be more engaged and provide a different learning style (Lesisko et al., 2010). The quality of teaching has been proven to be another factor in the level of student achievement (Lee et al., 2012). The overarching factor in all of these methods is assessment. Teachers can incorporate all these methods with quality interim assessment program (Wiliam & Black, 2010). TLI (2010) incorporates working with schools to align the curriculum, provide professional development for better teacher quality, and through technology provide instant feedback for teachers and students.

This research project was designed to provide additional research on the present limited amount concerning effectiveness of assessment method on mathematics achievement. The effects of a commercially available interim assessment program on students mathematics achievement was compared to locally made interim assessment program. TLI was used as the commercially available assessment program. The locally made interim assessment program was made from local teachers making the interim assessment and schedule of tests. The end of course geometry, end of course algebra I, and Eighth Grade Mathematics Benchmark test data were used for the comparison of mathematics achievement. As suggested by the research, teachers in either program were provided with professional development on the technology, curriculum alignment, and assessment design for successful development and implementation. (Chappuis et al., 2009; Helen, 2011; Wiliam & Black, 2009)

CHAPTER III

METHODOLOGY

A 2009 study conducted by Black and Wiliam reviewed how standards based assessment in the classroom can be raised to benefit the students and teachers. Assessment has been shown to be a key piece in raising the standards in the classroom by giving the teacher a measure of instruction effectiveness (Dunn & Mulvenon, 2009). The current research points to assessment as an integral feature in raising achievement. The development of assessments to measure student understanding of the standards during the course is essential (Black & Wiliam, 2009). Davis and McGowen (2009) stated that the concept of using commercially made interim assessments in classrooms, as an alternative to locally made assessments is a recent practice to raise student achievement. A review of literature illustrates that more research is needed in this area to help school leaders determine the best practices in the area of assessment. Two of the identified areas of assessments include research in use of commercially available assessments versus locally created assessments. This research project addressed examining at these two areas for mathematics at the high school level.

This study examined the effects of method of assessment and gender on algebra achievement for a population of Arkansas students after controlling for eighth grade mathematics achievement. Second, the study examined the effects of method of assessment and SES on algebra achievement for a population of Arkansas students after controlling for eighth grade mathematics achievement.

Third, the study examined the effects of method of assessment and gender on geometry achievement for a population of Arkansas students after controlling for algebra achievement. Finally, the purpose of this study was to determine the effects of method of assessment and SES on geometry achievement for a population of Arkansas students after controlling for algebra achievement.

The research hypotheses are as follows:

$H1_a$: There will be no significant interaction between type of assessment and gender on algebra achievement when controlling for eighth grade mathematics achievement.

$H1_b$: There will be no significant difference in the main effect of type of assessment on algebra achievement when controlling for eighth grade mathematics achievement.

$H1_c$: There will be no significant difference in the main effect of gender on algebra achievement when controlling for eighth grade mathematics achievement.

$H2_a$: There will be no significant interaction between type of assessment and SES on algebra achievement when controlling for eighth grade mathematics achievement.

$H2_b$: There will be no significant difference in the main effect of type of assessment on algebra achievement when controlling for eighth grade mathematics achievement.

$H2_c$: There will be no significant difference in the main effect of SES on algebra achievement when controlling for eighth grade mathematics achievement.

$H3_a$: There will be no significant interaction between type of assessment and gender on geometry achievement when controlling for algebra achievement.

$H3_b$: There will be no significant difference in the main effect of type of assessment on geometry achievement when controlling for algebra achievement.

$H3_c$: There will be no significant difference in the main effect of gender on geometry achievement when controlling for algebra achievement.

$H4_a$: There will be no significant interaction between type of assessment and SES on geometry achievement when controlling for algebra achievement.

$H4_b$: There will be no significant difference in the main effect of type of assessment on geometry achievement when controlling for algebra achievement.

$H4_c$: There will be no significant difference in the main effect of SES on geometry achievement when controlling for algebra achievement.

This chapter discusses the research design, the process of obtaining a sample, and a description of the sample population. The instrument used to measure student achievement is discussed and the data collection and statistical analysis processes is detailed. Finally, limitations of the study will be discussed.

Research Design

A quantitative, causal comparative strategy was used for this study. The independent variables for the Hypothesis 1 were method of assessment and gender. The dependent variable was algebra achievement. For these hypotheses, eighth grade mathematics achievement was used as a covariate. The independent variables for the Hypothesis 2 were method of assessment and gender. The dependent variable was geometry achievement. For these hypotheses, algebra achievement was used as a covariate. The independent variables for the Hypothesis 3 were method of assessment and SES. The dependent variable was algebra achievement. For these hypotheses, eighth grade mathematics achievement was again used as a covariate. Finally, the independent variables for the Hypothesis 4 were method of assessment and SES. The dependent variable was geometry achievement. For these hypotheses, algebra achievement was used as a covariate. The design of this study was used in order to isolate groups of students that used the same method of assessment in the initial year. The second year had a group change assessment methods. Using a covariate allowed for the two groups of students to be compared as if they started equally. The ex-post facto design was preferred because data was already available and could be gathered to show if a difference exists between methods of assessment. Weaknesses of this causal-comparative research design are that experimental controls or variables cannot be manipulated since they have already occurred. In addition, caution must be applied in interpreting results due to groups have already been previously assigned (Lord, 1973).

Sample

The sample for this study consisted of 711 students who took the end of course geometry and algebra exams in six Arkansas schools in the River Valley area of the state. Three schools where chosen based on their initial participation in using TLI during the 2009-2010 school year. Although these schools had used other traditional methods of assessment that included those developed by their local district during the 2008-2009 school year, they had switched to TLI during the 2009-2010 school year. A comparison sample was also drawn from three different schools in the Arkansas River Valley that used traditional methods of assessment developed by their local district during the 2008-2009 and 2009-2010 school years. All the schools from which samples were drawn shared similar demographic characteristics. These demographics included student to teacher ratio, percentage of free or reduced lunch, ethnic makeup, and gender percentages. All students sampled for this study were traditional 9th graders when they took algebra I and were also traditional 10th graders when they took geometry. The students were chosen based on convenience techniques from the total population in each school that met the above criteria and with the exclusions listed in the next section.

Exclusion Criteria

Students who did not complete the ACTAAP end of course exams in the 2009-2010 school years were excluded from the study. Students not having a corresponding ACTAAP exam from the same school in the previous school year were also eliminated. Any student that was repeating the grade as a non-traditional 9th grader in algebra I or 10th grader in geometry was not included in the study.

Demographics

All algebra I courses in 2009-2010 consisted of students that were in various grades and some were students that were repeating the course due to past failures. In the requested sample, there were 362 students; of the 362 students, 176 boys and 186 girls were enrolled as a traditional ninth grader in algebra I for the first time. These same students took the eighth grade mathematics benchmark in their previous year. Table 1 shows the demographic breakdown for the students in the algebra I population.

Table 1

Demographics for 2009-2010 Algebra I Students

Characteristic	TLI *n* =196 (%)		Local Assessment *n* = 166 (%)	
Race				
White	154	(78.6)	136	(81.9)
African American	5	(2.6)	3	(1.8)
Hispanic	33	(16.8)	22	(13.3)
Asian Pacific Islander	3	(1.5)	4	(2.4)
Native American	1	(0.5)	1	(0.6)
Gender				
Female	92	(46.9)	94	(56.6)
Male	104	(53.1)	72	(43.4)
SES				
Economic Disadvantaged	121	(61.7)	76	(45.8)
Not Economic Disadvantaged	75	(38.3)	90	(54.2)

Table 1 shows the breakdown for students in the 2009-2010 algebra I population. This shows the number of students by gender that used TLI versus those that did not in 2009-2010. Approximately 49% of the girls and 59% of the boys in the algebra I sample came from a school that used TLI in 2009-2010. In the alternate case, 46% of the students came from a school that used locally made assessments during the 2009-2010 school year. In the algebra I class, there were 362 students; of the 362 students, 121 economic disadvantaged students and 75 not economic disadvantaged students were enrolled in schools that used TLI in 2009-2010. Schools that used locally made assessments in 2009-2010 consisted 76 economic disadvantaged and 90 not economic disadvantaged students. Table 3 shows the demographic breakdown for the students in by SES and assessment type.

All geometry courses in 2009-2010 consisted of students that were in various grades and some were students that were repeating the course due to past failures. In the requested sample, there were 349 students; of the 349 students, 172 boys and 177 girls were enrolled as a traditional 10th grader in geometry for the first time. These same students took the algebra end of course assessment in their previous year. Table 4 shows the demographic breakdown for the students in the geometry population.

Table 2

Demographics for 2009-2010 Geometry Students

Characteristic	TLI *n* =195(%)		Local Assessment *n* = 154(%)	
Race				
White	155	(79.5)	126	(81.8)
African American	4	(2.0)	0	(0.0)
Hispanic	30	(15.5)	23	(14.9)
Asian Pacific Islander	4	(2.0)	5	(3.3)
Native American	2	(1.0)	0	(0.0)
Gender				
Female	97	(46.9)	80	(56.6)
Male	98	(53.1)	74	(43.4)
SES				
Economic Disadvantaged	106	(54.4)	73	(47.4)
Not Economic Disadvantaged	89	(45.6)	81	(52.6)

Table 2 shows the breakdown for students in the 2009-2010 geometry population. This shows the number of students by gender that used TLI versus those that did not in 2009-2010. Approximately 55% of the girls and 57% of the boys in the algebra I sample came from a school that used TLI in 2009-2010. In the alternate case, 56% of the students came from a school that used locally made assessments during the 2009-2010 school year. In the geometry class, there were 349 students; of the 349 students, 59.2% economic disadvantaged students and 52.3% not economic disadvantaged students were enrolled in schools that used TLI in 2009-2010. Schools that used locally made assessments in 2009-2010 consisted 40.8% economic disadvantaged and 47.7% not economic disadvantaged students. Table 6 shows the demographic breakdown for the students in by SES and assessment type.

Instrumentation

The ACTAAP end of course examination was the instrument for this study. According to the Arkansas Department of Education (2010), The ACTAAP includes an End-of-Course Examination in geometry and algebra I. It consists of multiple-choice and open-response questions that directly assess student knowledge based on the Arkansas Geometry and Algebra I Mathematics Curriculum Framework. The geometry exam consists of eight sessions over a 2-day period. The first session consists of 20 multiple-choice questions. The second session consists of 15 multiple-choice questions. The third session consists of three open response questions. The fourth session consists of 20 multiple-choice questions. The fifth session consists of 15 multiple-choice questions. The sixth session consists of two open response questions. The seventh session consists of 20 multiple-choice questions. The final session consists of two open response questions.

The algebra I exam consists of eight sessions over a 2-day period. The first session consists of 20 multiple-choice questions (Arkansas Department of Education, 2010). The second session consists of 15 multiple-choice questions. The third session consists of three open response questions. The fourth session consists of 20 multiple-choice questions. The fifth session consists of 15 multiple-choice questions. The sixth session consists of two open response questions. The seventh session consists of 20 multiple-choice questions. The final session consists of two open response questions.

ACTAAP end of course geometry of 2009-2010 reliability was assessed through internal-consistency measures and inter-rater reliability (Pearson Education, 2010). The ACTAAP end of course geometry used evidence internal structure and evidence of fairness (e.g. differential item functioning) to support the validity of the ACTAAP end of course geometry assessment. The measures for Cronbach's alpha are 0.922. These coefficients indicate a strong internal consistency. The measure for Inter-rater reliability is 99.2% for agreement (Pearson Education, 2010).

ACTAAP end of course algebra I of 2009-2010 reliability was assessed through internal-consistency measures and inter-rater reliability (Pearson Education, 2010). The ACTAAP End of algebra I used evidence internal structure and evidence of fairness (e.g. differential item functioning) to support the validity of the ACTAAP end of course algebra I assessment. The measures for Cronbach's alpha are 0.914. These coefficients indicate a strong internal consistency. The measure for Inter-rater reliability is 99.1% for agreement (Pearson Education, 2010).

ACTAAP end of course algebra I of 2008-2009 reliability was assessed through internal-consistency measures and inter-rater reliability (Pearson Education, 2009). The ACTAAP End of algebra I used evidence internal structure and evidence of fairness (e.g. differential item functioning) to support the validity of the ACTAAP end of course algebra I assessment. The measures for Cronbach's alpha are 0.920. These coefficients indicate a strong internal consistency. The measure for Inter-rater reliability is 98.2% for agreement (Pearson Education, 2009).

ACTAAP eighth grade mathematics benchmark assessment reliability was assessed through internal-consistency measures and inter-rater reliability (Pearson Education, 2009). The ACTAAP eighth grade mathematics benchmark assessment used evidence internal structure and evidence of fairness (e.g. differential item functioning) to support the validity of the ACTAAP eighth grade mathematics benchmark assessment. The measures for accuracy are (.92). The measures for consistency are (.89). These coefficients indicate a strong internal consistency. The measure for Inter-rater reliability is 99% for average agreement (Pearson Education, 2009).

Data Collection Procedures

Following Institutional Review Board approval on July 18, 2011 (see Appendix), the local district collected with help from the Arkansas Research Center without personal identifying information. The data were requested of those schools that participated in TLI in the 2009-2010 school year or not participating and using locally made assessments. The researcher requested the demographic data that included gender as male or female and SES as either economic disadvantage or not economic disadvantage. The socioeconomic indicator is based on free/reduced lunch status as provided by the school. The scale score and proficiency status was requested for students that participated in the end of course algebra assessment in the 2009-2010 school year and the previous year scale score for eighth grade mathematics benchmark. The scale score and proficiency status was also requested for students that participated in the end of course geometry assessment in the 2009-2010 school year and the previous year scale score for end of course algebra exam. The data were received via email during the fall of 2011. Only students who completed both the 2009 and 2010 ACTAAP testing seasons at the same school were included in the samples that were requested. Data were stored on two USB drives. The drives were kept locked in a fireproof safe when not being used by the researcher.

Analytical Methods

To address the first group of hypotheses ($H_{1a, b, c}$), a 2 x 2 factorial analysis of covariance was conducted to examine the interactions and main effects of method of assessment and gender as the independent variables and the algebra achievement as the dependent variable and eighth grade achievement as the covariate. To address the next three hypotheses ($H_{2a, b, c}$) , a 2 x 2 factorial analysis of covariance was conducted examine the interactions and main effects of type of assessment and gender as the independent variables and the geometry achievement as the dependent variable and algebra achievement as the covariate. To address the third set of hypotheses ($H_{3a, b, c}$), a 2 x 2 factorial analysis of covariance was conducted examine the interactions and main effects of type of assessment and SES as the independent variables and the algebra achievement as the dependent variable and eighth grade achievement as the covariate. Finally, to address the last three hypotheses ($H_{4a, b, c}$), a 2 x 2 factorial analysis of covariance was conducted examine the interactions and main effects of type of assessment and SES as the independent variables and the geometry achievement as the dependent variable and algebra achievement as the covariate A two-tailed test with a .05 level of significance was used for all the tests in this study.

Limitations

It is important to note any limitations that might have an adverse effect on the results of this study. This allows the reader to determine what if any effect these conditions might have had upon the interpretation of the results. The following were limitations associated with this study.

The first limitation was that there were multiple schools that were used in the study. This may lead to varying degrees of implementation of practices outlined by TLI. The variances in implementation could impact the construct validity by impacting mathematics achievement. This may have little impact on internal validity, but could affect external validity due to difficulty replicating the experiment. Another limitation was that the locally made assessments could also vary due to the multiple districts involved in the study. This could influence construct validity because students may have variation in mathematics achievement. This may have little impact on internal validity, but could impact external validity due to difficulty replicating the experiment.

A third limitation was that the sample of students was limited due to the constraints of having a school that made the transition to TLI during this timeframe of the study and used locally made assessments in the previous year. The small sample size could have an impact on construct validity by skewing mathematics achievement. This may have little impact on internal validity, but could affect external validity due to difficulty replicating the experiment. A fourth limitation was that only high school mathematics courses were considered. Commercially made assessments were also available for literacy and science courses. This could also influence construct validity by skewing mathematics achievement by using a small sample of exams. This may have little impact on internal validity, but could impact external validity due to difficulty replicating the experiment.

CHAPTER IV

RESULTS

A quantitative approach was adopted in this study to examine the effects of method of assessment by gender on algebra achievement for a sample of Arkansas students while controlling for eighth grade mathematics achievement. This study was also aimed at determining the effects of method of assessment by SES on algebra achievement for a sample of Arkansas students while controlling for eighth grade mathematics achievement. Yet, another purpose of this study was to determine the effects of method of assessment by gender on geometry achievement for a sample of Arkansas students while controlling for algebra achievement. Finally, this study sought to determine the effects of method by SES on geometry achievement for a sample of Arkansas students while controlling for algebra achievement.

The independent variables for the first set of hypotheses were method of assessment and gender. The dependent variable was algebra achievement. For these hypotheses, eighth grade mathematics achievement was used as a covariate. The independent variables for the second set of hypotheses were method of assessment and SES. The dependent variable was algebra achievement. For these hypotheses, eighth grade mathematics achievement was used as a covariate. The independent variables for the third set of hypotheses were method of assessment and gender. The dependent variable was geometry achievement. For these hypotheses, algebra achievement was used as a covariate. The independent variables for the fourth set of hypotheses were method of assessment and SES. The dependent variable was geometry achievement. For these hypotheses, algebra achievement was used as a covariate. Factorial Analysis of Covariance (ANCOVAs) were run to test at each of the study's null hypotheses. However, concerning Hypotheses 4a, 4b and 4c, a major violation of the assumptions for ANCOVA led to an adjustment of the analysis from a factorial ANOCOVA to a factorial ANOVA. Prior to conducting data analysis, scatterplots were examined to test the assumptions of linearity between the covariate and dependent variable. These assumptions were examined and met for Hypotheses 1 and 2 ($r^2 = .613$), as well as for Hypotheses 3 and 4 ($r^2 = .655$).

Hypothesis 1

Hypothesis 1_a stated there will be no significant interaction between type of assessment and gender on algebra achievement when controlling for eighth grade mathematics achievement. Hypothesis 1_b stated there will be no significant difference in the main effect of type of assessment on algebra achievement when controlling for eighth grade mathematics achievement while Hypothesis 1_c stated there will be no significant difference in the main effect of gender on algebra achievement when controlling for eighth grade mathematics achievement. Before conducting an ANCOVA, preliminary analysis was conducted to estimate if the distribution of algebra achievement in the populations from which the samples were drawn was relatively normal. An examination of box and whisker plots revealed no significant outliers among the groups. Kolmogorov-Smirnov tests further confirmed that the distribution of algebra 1 scores across all groups could be assumed to be normal (male non-TLI students p = .200; female non-TLI p = .200; male TLI students p = .171; female TLI students p = .200).

To test the assumption of homogeneity of variances, Levene's test was conducted and determined to be significant $F(3,358) = 5.41$, $p = .001$, therefore the assumption of homogeneity of variances was violated. However, this violation was not deemed critical such as would require any adjustment in the method of data analysis. Finally, results of the assumption of homogeneity of regression slopes yielded non-significant results $F(3, 355) = 1.19$, $p = .310$; therefore this assumption was met. Once preliminary data analysis was completed, an ANCOVA to test Null Hypotheses 1_a, 1_b, and 1_c was conducted. Table 3 provides a summary of the descriptive statistics related to this analysis.

Table 3

Mean Algebra 1 Scores by Assessment Type and Gender Using Eighth grade Mathematics Benchmark Scores as a Covariate

	Unadjusted			*Adjusted*	
	N	*M*	*SD*	*M*	*SE*
TLI Females	92	224.16	35.68	231.02	2.52
TLI Total	196	219.57	40.62	228.40	1.76
Non-TLI Male	72	251.64	26.29	240.17	2.88
Non-TLI Female	94	247.55	34.19	238.27	2.51
Non-TLI Total	166	249.33	30.99	239.22	1.94

The test results revealed that the covariate (eighth grade mathematics scores) was statistically significant, $F(1, 357) = 468.28$, $p < .001$, $\eta^2 = .567$. However, the interaction between gender and instruction was not statistically significant, $F(1, 357) = 1.96$, $p = .163$. Furthermore, the main effect for gender was not statistically significant, $F(1, 357) = 0.43$, $p = .512$; but the main effect for assessment type was statistically significant, $F(1, 357) = 16.14$, $p < .001$, $\eta^2 = 043$. (See Table 4)

Table 4

Analysis of Covariance for Mathematics Achievement (Algebra 1) as a Function of Assessment Type and Gender, After Controlling for Eighth Grade Mathematics Benchmark Scores

Source	*SS*	*df*	*MS*	*F*	*Sig.*	*ES*
8th Mathematics	270044.62	1	270044.62	468.28	.000	.567
Gender	247.87	1	247.87	0.43	.512	.001
Assessment	9306.94	1	9306.94	16.14	.000	.043
Gender*Assessment	1129.40	1	1129.40	1.96	.163	.005
Error	205871.77	357	576.67			
Total	20248804.00	362				

Figure 1 provides a visual summary of mean algebra performance across the different groups. Based on these results, Null Hypotheses 1_a and 1_c could not be rejected; however, Null Hypothesis 1_b was rejected. This means that the covariate eighth grade mathematics score significantly adjusted the effects of the two factors; however, gender and instruction did not work together to affect mathematics achievement after controlling for eighth grade mathematics achievement. Furthermore, the results also suggest that assessment on its own significantly impacted mathematics achievement, while gender did not appear to have a similar independent effect.

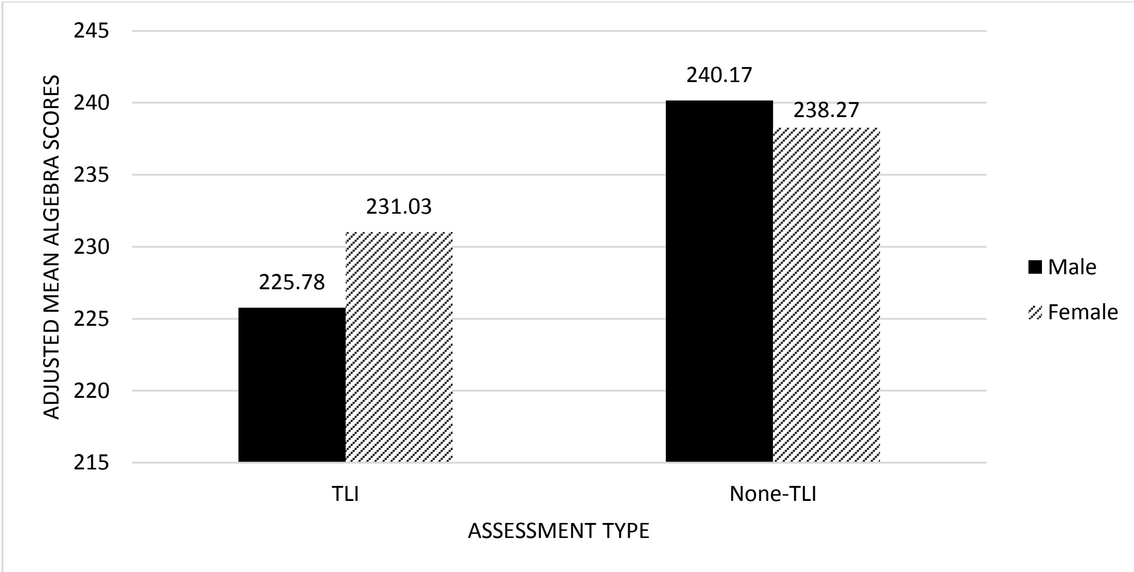

Figure 1. Adjusted mean algebra 1 scores by assessment type and gender.

In other words, the algebra achievement for the non-TLI students (*M* = 239.22, *SE* = 1.94), was significantly higher than that of the TLI students (*M* = 228.40, *SE* = 1.76). Conversely, the difference in algebra 1 scores between males and females was such as could be attributed to measurement or sample errors; and did not represent a true mean difference in the population.

Hypothesis 2

Hypothesis 2_a stated there will be no significant interaction between type of assessment and SES on algebra achievement when controlling for eighth grade mathematics achievement. Hypothesis 2_b stated there will be no significant difference in the main effect of type of assessment on algebra achievement when controlling for eighth grade mathematics achievement while Hypothesis 2_c stated there will be no significant difference in the main effect of SES on algebra achievement when controlling for eighth grade mathematics achievement. Before conducting an ANCOVA, preliminary analysis was conducted to assess if the distribution of algebra achievement in the populations from which the samples were taken could be assumed to be normal. An examination of box and whisker plots revealed no major outlier within any of the groups. Kolmogorov-Smirnov tests indicated that the distribution of algebra scores for all groups could be assumed to be normal (non-economic disadvantaged non -TLI students $p = .200$; economic disadvantaged non TLI $p = .200$; non-economic disadvantaged TLI students $p = .200$; economic disadvantaged TLI students $p = .200$).

To test the assumption of homogeneity of variances, Levene's test was conducted and determined to be significant $F(3, 358) = 4.73$, $p = .003$, therefore the assumption of homogeneity of variances was violated for this analysis. However, given that the groups were all relatively large and somewhat similar size, this violation was not considered critical such as would merit a change or adjustment in the analysis. Finally, results of the assumption of homogeneity of regression slopes yielded non-significant results $F(3, 355) = 1.03$, $p = .38$ indicating that this assumption was met. Once preliminary data analysis was completed, an ANCOVA to test Null Hypotheses 2_a, 2_b, and 2_c was conducted. Table 5 provides a summary of descriptive statistics.

Table 5

Mean Algebra 1 Scores by Assessment Type and SES Using Eighth grade Mathematics Benchmark Scores as a Covariate

	Unadjusted			*Adjusted*	
	N	*M*	*SD*	*M*	*SE*
TLI Non-Economic	75	227.71	43.64	228.30	2.78
TLI Economic	121	214.53	37.95	228.16	2.28
TLI Total	196	219.57	40.62	228.32	1.80
Non-TLI/Non-Economic	90	255.68	28.04	240.99	2.63
Non-TLI/Economic	76	241.80	32.78	236.91	2.77
Non-TLI Total	166	249.33	30.99	238.95	1.93

The test results revealed that the covariate (eighth grade mathematics scores) was statistically significant, $F(1, 357) = 444.96$, $p < .001$, $\eta^2 = .555$. However, the interaction between SES and instruction was not statistically significant, $F(1, 357) = 0.59$, $p = .445$. Furthermore, the main effect for SES was not statistically significant, $F(1, 357) = 0.64$, $p = .423$; but the main effect for assessment type was statistically significant, $F(1, 357) = 15.77$, $p < .001$, $\eta^2 = 042$ (See Table 6).

Table 6

Analysis of Covariance for Mathematics Achievement as a Function of SES, Using Eighth Grade Mathematics Benchmark Scores as a Covariate

Source	*SS*	*df*	*MS*	*F*	*Sig.*	*ES*
8th Mathematics	257601.16	1	257601.16	444.96	.000	.555
SES	372.41	1	372.41	0.64	.423	.002
Assessment	9129.70	1	9129.70	15.77	.000	.042
SES*Assessment	338.86.40	1	338.86	0.59	.445	.002
Error	206678.23	357	578.93			
Total	20248804.00	362				

Figure 2 provides a visual summary of the relationship between the different groups in regards to algebra performance. Based on these results, Null Hypotheses 2_a and 2_c could not be rejected; however, Hypothesis 2_b was rejected. This means that the covariate eighth grade mathematics score significantly adjusted the effects of the two factors; however, SES and instruction did not work together to affect mathematics achievement after controlling for eighth grade mathematics achievement. Furthermore, the effect of assessment was statistically significant while that of SES was not statistically significant.

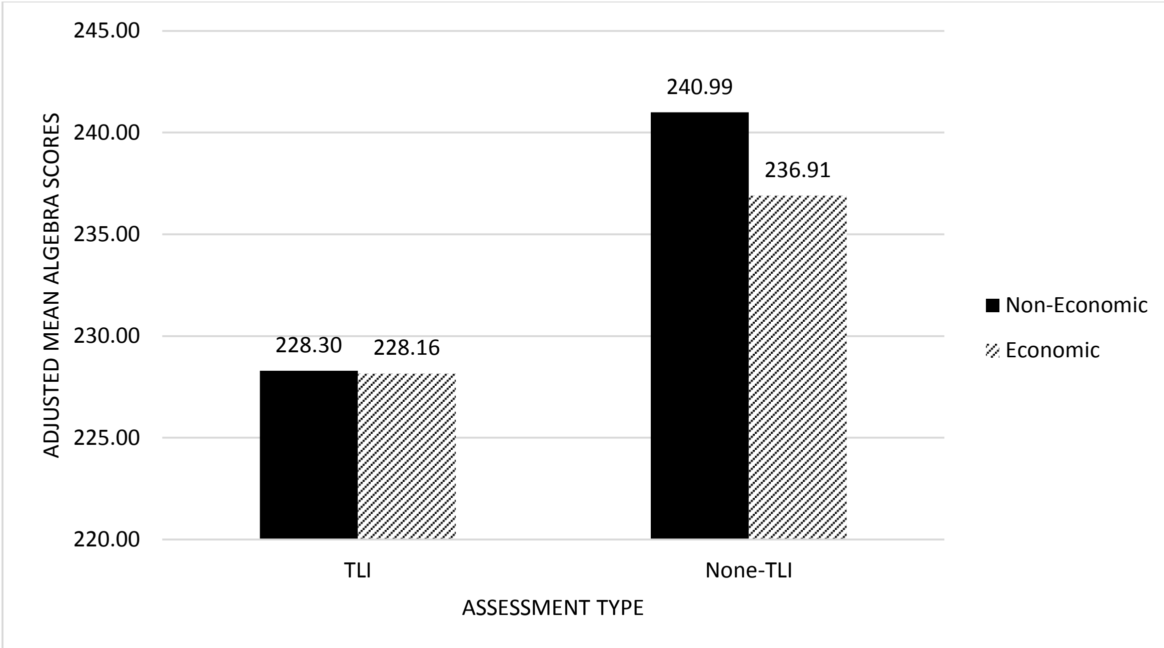

Figure 2. Adjusted mean algebra 1 scores by SES and assessment type.

Once again, the results show that the adjusted mean algebra 1 performance of the TLI students (M = 228.32, SE = 1.80) to be significantly lower than those of the non-TLI students (M = 238.95, SE = 1.93). At the same time, the observed difference between students based on SES was negligible enough as to be attributed to sampling or measurement errors rather than a true difference in the relevant populations.

Hypothesis 3

Hypothesis 3_a stated there will be no significant interaction between type of assessment and gender on geometry achievement when controlling for algebra mathematics achievement. Hypothesis 3_b stated there will be no significant difference in the main effect of type of assessment on geometry achievement when controlling for algebra mathematics achievement while Hypothesis 3_c stated there will be no significant difference in the main effect of gender on geometry achievement when controlling for algebra mathematics achievement. Before conducting an ANCOVA, preliminary analysis of the data included checks for outliers, normality, homogeneity of variances, and homogeneity of regression slopes. Box and whisker plots were used to check for outliers. This check did not reveal any outlier worth noting. Kolmogorov-Smirnov tests also indicated that the assumption of normality was met for all but one of the groups (male non-TLI students $p = .037$; female non-TLI $p = .200$; male TLI students $p = .200$; female TLI students $p = .106$). Again, because of the relatively large sample size for each group, this violation was not deemed critical, as ANOVA is considered robust to minor violations of this assumption (Mertler & Vannatta, 2013). To test the assumption of homogeneity of variances, Levene's test was conducted and determined to be non-significant $F(3, 345) = 0.62$, $p = .604$; therefore, the assumption of homogeneity of variances was not violated. Finally, results of the assumption of homogeneity of regression slopes yielded non-significant results $F(3, 342) = 1.57$, $p = .196$, indicating that this assumption was met. Once preliminary data analysis was completed, an ANCOVA to test Null Hypotheses 3_a, 3_b, and 3_c was conducted. Table 7 provides a summary of the descriptive statistics for this analysis.

Table 7

Mean Geometry Scores by Assessment Type and Gender Using Algebra 1 Scores as a Covariate

	Unadjusted			*Adjusted*	
	N	*M*	*SD*	*M*	*SE*
TLI Males	98	216.72	31.65	221.16	2.05
TLI Females	97	209.80	34.44	251.08	2.06
TLI Total	195	213.28	33.16	218.12	1.46
Non-TLI Male	74	233.81	32.54	227.27	2.36
Non-TLI - Female	80	229.45	36.41	223.68	2.27
Non-TLI Total	154	231.55	34.56	225.48	1.65

The test results revealed that the covariate (algebra 1 scores) was statistically significant, $F(1, 344) = 621.10$, $p < .001$, $\eta^2 = .644$. However, the interaction between gender and instruction was not statistically significant, $F(1, 344) = 0.33$, $p = .568$. Furthermore, the main effect for gender was statistically significant, $F(1, 344) = 4.93$, $p = .027$, $\eta^2 = .014$; as well as the main effect for assessment type $F(1, 344) = 10.98$, $p = .001$, $\eta^2 = 031$. (See Table 8)

Table 8

Analysis of Covariance for Mathematics Achievement (Geometry) as a Function of Assessment Type and Gender, after Controlling for Algebra 1

Source	*SS*	*df*	*MS*	*F*	*Sig.*	*ES*
Algebra	252905.02	1	252905.02	174.78	.000	.670
Gender	2008.25	1	2008.25	4.93	.027	.014
Assessment	4471.51	1	4471.51	10.98	.000	.031
Gender*Assessment	132.99	1	132.99	0.33	.568	.001
Error	140072.98	344	407.19			
Total	17522892.00	349				

Figure 3 provides a visual summary the adjusted mean differences in mathematics achievement (geometry) achievement between the groups. Based on these results, Null Hypothesis 3_a could not be rejected; however, Null Hypotheses 3_b and 3_c were rejected. This means that the covariate algebra 1 mathematics score significantly adjusted the effects of the two factors. Gender and instruction however, did not work together to affect geometry achievement after controlling for algebra 1 achievement. Despite the absence of interaction between the two independent variables, their independent effects were statistically significant. This means that the difference in geometry achievement between TLI students (*M* =218.12, *SE* = 1.46) was significantly lower than that of Non-TLI students (*M* = 225.48, *SE* = 1.65).

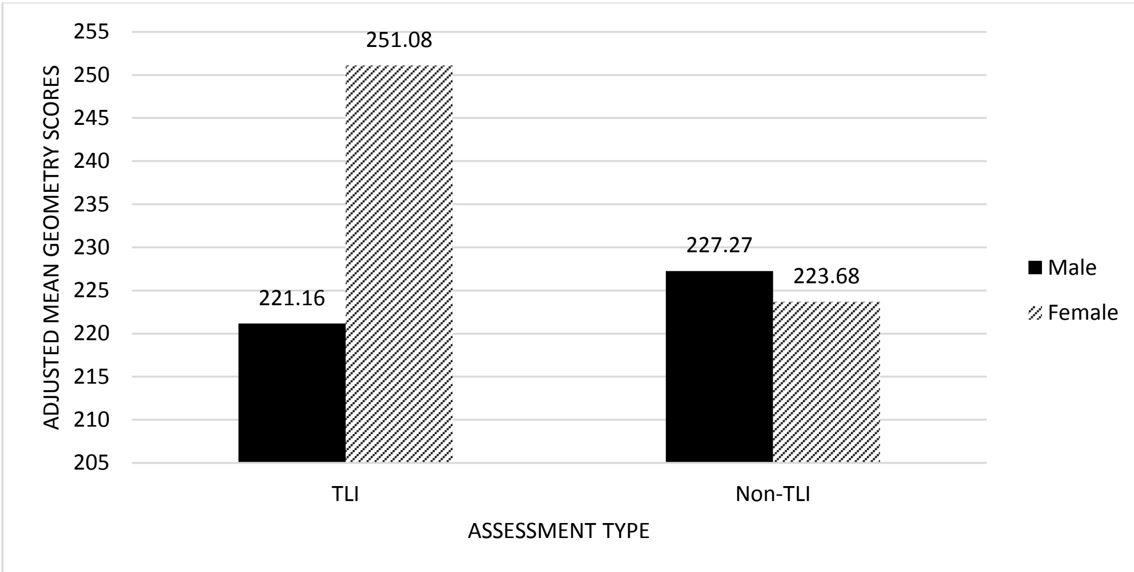

Figure 3. Adjusted mean geometry scores by assessment type and gender.

Similarly, the adjusted mean differences in geometry achievement of males students (*M* = 224.21, *SE* = 1.55) was significantly higher than that females students (*M* = 219.38, *SE* = 1.52) when considered independent of assessment type. The implication here being that these differences concerning both variables, were large enough to considered true differences in the relevant populations.

Hypothesis 4

Hypothesis 4_a stated there will be no significant interaction between type of assessment and SES on geometry achievement when controlling for algebra mathematics achievement. Hypothesis 4_b stated there will be no significant difference in the main effect of type of assessment on geometry achievement when controlling for algebra mathematics achievement while Hypothesis 4_c stated there will be no significant difference in the main effect of SES on geometry achievement when controlling for algebra mathematics achievement. Before conducting an ANCOVA, preliminary analysis was conducted to examine the relevant assumptions. An examination of box and whisker plots revealed no significant outliers among the groups. Kolmogorov-Smirnov tests indicated that the population distribution of algebra scores for all groups could be assumed to be normal (non-economic disadvantaged non-TLI students p = .200; economic disadvantaged non TLI p = .200; non-economic disadvantaged TLI students p = .200; economic disadvantaged TLI students p = .200). Levene's test was not significant F(3, 345) = 1.21, p = .305, therefore the assumption of homogeneity of variances was not violated. However, results of a preliminary ANCOVA revealed that an assumption of homogeneity of regression slopes was not tenable F(3, 342) = 3.06, p = .028. The violation of this critical assumption for ANCOVA is known to complicate the interpretation of test results (Mertler & Vannatta, 2013). To avoid such complications, geometry 1 was dropped as a covariate in the analysis of Null Hypothesis 4. Furthermore, a two-way factorial ANOVA was used in place of a two-way factorial ANCOVA to test Null Hypotheses 4_a, 4_b, and 4_c. Table 9 provides a summary of the descriptive statistics for this analysis.

Table 9

Mean Geometry Scores by Assessment Type and SES

	N	*M*	*SD*
TLI Non-Economic	89	219.51	32.13
TLI Economic	106	208.06	33.26
TLI Total	195	213.28	33.16
Non-TLI Non-Economic	81	243.07	29.62
Non-TLI - Economic	73	218.75	35.32
Non-TLI Total	154	231.55	34.56

Results of the two-way ANOVA revealed that the interaction between SES and instruction was not statistically significant, $F(1, 345) = 3.33$, $p = .069$. Furthermore, the main effect for SES was statistically significant, $F(1, 345) = 25.75$, $p < .001$, $\eta^2 = .069$; and the main effect for assessment type was statistically significant, $F(1, 345) = 23.63$, $p < .001$, $\eta^2 = .064$. (See Table 10)

Table 10

Analysis of Variance for Mathematics Achievement (Geometry) as a Function of Assessment Type and SES

Source	*SS*	*df*	*MS*	*F*	*Sig.*	*ES*
SES	27389.28	1	27389.28	25.75	.000	.069
Assessment	25133.87	1	25133.87	23.63	.000	.064
SES*Assessment	3546.65	1	3546.65	3.33	.069	.010
Error	366991.03	345	1063.74			
Total	17522892.00	349				

Figure 4 provides a visual summary of the unadjusted mean geometry achievement scores across the different groups. Based on these results, Null Hypothesis 4_a was not rejected; however, Null Hypotheses 4_b and 4_c were rejected. This means that although SES and instruction did not work together to affect mathematics achievement; independently, both assessment type and SES appeared to have an effect on mathematics (geometry) achievement. The results once again provided evidence to show that the mean performance of TLI students (*M* = 213.78, *SE* = 2.35) was significantly lower than that of Non-TLI students (*M* = 230.91, *SE* = 2.63).

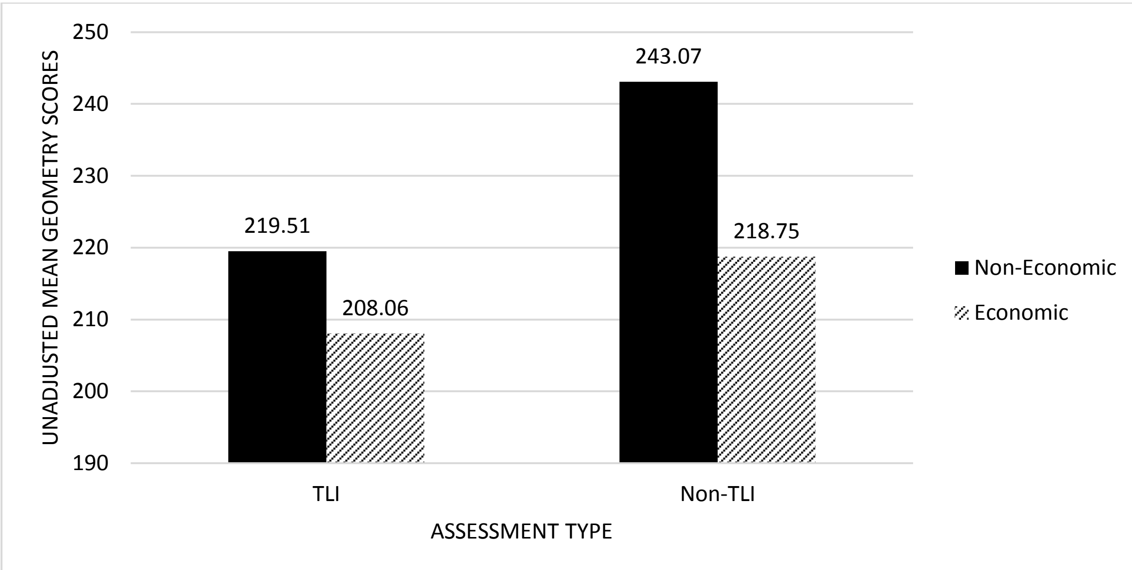

Figure 4. Unadjusted mean geometry scores by assessment type and SES.

Similarly, the mean geometry achievement scores for non-economic disadvantages students (*M* = 231.29, *SE* = 2.50) was significantly higher than the mean geometry scores for the economic disadvantaged students (*M* = 213.41, *SE* = 2.48). This implication here being that these differences were large enough for support the conclusion that they represented true differences in the population and not chance differences due to sampling error.

CHAPTER V

DISCUSSION

School leaders are always searching for tools that will help teachers increase academic achievement. One such tool that is used to measure student achievement is assessment. Many types of assessment exist and educators should survey all available research about these assessments that will help with student achievement. The objective of this study was to contribute to the body of research in determining the academic effectiveness of interim assessment in the high school setting.

Specifically, it was the purpose of this study to examine the effects of assessment type on students' mathematics achievement by gender or SES. For this purpose, a causal-comparative study was designed using students drawn from six Arkansas schools; where students either took the end of course algebra I exam as a 9th grade student or the end of course geometry exam as a 10th grade student. In addition to data on each participant's mathematics achievement (algebra and geometry), data were also obtained on their gender, SES, and previous mathematics achievement.

This chapter includes a summary of conclusions based on the results in this study. Following the conclusion, recommendations based on the conclusions are presented. These recommendations include practical suggestions for school administrators, as well as ideas for consideration by policymakers. Finally, the implications and significance of the study are discussed.

Conclusions

The analysis of data in the previous chapter led to several conclusions. The conclusion for each hypothesis will be discussed followed by a summary of conclusions related to the overall purpose of the study.

Hypothesis 1

Hypothesis 1_a stated there would be no significant interaction between type of assessment and gender on algebra achievement when controlling for eighth grade mathematics achievement. Hypothesis 1_b stated there would be no significant difference in the main effect of type of assessment on algebra achievement when controlling for eighth grade mathematics achievement. Hypothesis 1_c stated there would be no significant difference in the main effect of gender on algebra achievement when controlling for eighth grade mathematics achievement. There was no significant interaction between the independent variables gender and assessment type on the dependent variable mathematics achievement measured by the algebra I end of course exam. Gender and assessment type did not work together as factors to influence algebra achievement. For the main effect of assessment type, a significant difference in algebra achievement however was found between students in the Non-TLI assessment groups and those in the TLI assessment group. However, a significant difference was not found in algebra achievement for students based on gender.

Students that used Non-TLI assessments had higher mean scores than their TLI counter parts. These results mirror those of Davis and McGowen (2009) who stated that there are limitations to what commercial programs can do concerning student achievement. The findings here are however quite different from those of (Dunn &

Mulvenon, 2009), who indicated that there was ample evidence to support the use of commercial assessments to improve student achievement. One possible explanation for this disparity may be the level of buy-in of teachers and other stakeholders. According to Shepard (2009), the buy-in of the teachers who are heavily involved in curriculum and assessment development is critical to the successful implementation of instructional interventions. Such stakeholders may be more resistant to commercial assessment programs (in contrast to locally developed programs) which are often bought and implemented without their direct input in the design. This was the case in an Australian study where Brown and Hirchfeld (2007) found that student perception of teacher buy-in of an assessment directly impacted student achievement. Although this was not directly examined in the current study, such factors may have had some impact on the results.

Hypothesis 2

Hypothesis 2_a stated there would be no significant interaction between type of assessment and SES on algebra achievement when controlling for eighth grade mathematics achievement. Hypothesis 2_b stated there would be no significant difference in the main effect of type of assessment on algebra achievement when controlling for eighth grade mathematics achievement. Hypothesis 2_c stated there would be no significant difference in the main effect of socioeconomic on algebra achievement when controlling for eighth grade mathematics achievement. There was no significant interaction between the independent variables SES and assessment type on the dependent variable mathematics achievement measured by the algebra I end of course exam. SES and assessment type did not work together as a factor to influence algebra achievement.

For the main effect of assessment type, a significant difference in algebra achievement was seen between students in the Non-TLI assessment group and those in the TLI assessment group. However, a significant difference was not found in algebra achievement for students based on SES. In other words, students that used Non-TLI assessments had higher mean scores than their TLI counter parts, while SES did not appear to have an impact on algebra achievement.

Again, these differences based on assessment type are contrary to what would be expected based on the findings of Dunn and Mulvenon (2009) among others. However, here again as suggested by Shepard (2009) and Brown and Hirschfeld (2007), this difference may be attributable to factors such as buy-in of the teachers. The findings with regard to SES were also surprising giving the wealth of evidence in the literature that suggests the great influence of socioeconomic factors on student achievement (Gasbarra & Johnson, 2008; Myers, 1986; Rutherford et al., 2010).

Hypothesis 3

Hypothesis 3_a stated there would be no significant interaction between type of assessment and gender on geometry achievement when controlling for algebra mathematics achievement. Hypothesis 3_b stated there would be no significant difference in the main effect of type of assessment on geometry achievement when controlling for algebra mathematics achievement. Hypothesis 3_c stated there would be no significant difference in the main effect of gender on geometry achievement when controlling for algebra mathematics achievement. There was no significant interaction between the independent variables gender and assessment type on the dependent variable mathematics achievement measured by the geometry end of course exam. Gender and assessment type did not work together as a factor to influence geometry achievement. For the main effect of assessment type, a significant difference in geometry achievement was seen between students in the Non-TLI assessment group and those in the TLI assessment group. A significant difference was also found in geometry achievement for students based on gender such that female students outperformed the male students in geometry.

In regards to geometry achievement, the difference found between assessment type appear to be in line with the study of Davis and McGowen (2009) which resulted in limited improvement in student achievement with commercial programs. However, these would not be expected based on the findings of Dunn and Mulvenon (2009) that showed that commercial programs positively impacted student achievement. Furthermore, the results showed that female students had higher geometry achievement than their male counterparts. These findings do not correspond with those of similar studies in the literature that shows that males outperform females in mathematics as first year engineering students (Rutherford et al., 2010). Similarly, Wilson and Zhang (1998) also found that in grades three and eight, males scored higher than females on the mathematics portion of the Iowa Test of Basic Skills. Heinrich et al. (2009) noted that female students that scored high in middle and high school mathematics standardized tests were also are enrolled in supplemental education services. Supplemental education services are a type of tutoring service that aids students with learning deficiencies. In the current study, there was no way of determining if the female students were enrolled in such programs, and the male students were not. However, considering a typical Arkansas public school setting, it would be safe to assume that this was not the case. So how do we make sense of the reversed direction of the gender difference in geometry achievement found in this study? A possible contributing factor, according to Brown and Hirschfeld (2007), may be that the disproportionately larger number of females in the sample compared to males, coupled with the relatively small overall sample size. Whatever the case may be, a further investigation of this phenomenon is necessary.

Hypothesis 4

Hypothesis 4_a stated there would be no significant interaction between type of assessment and SES on geometry achievement when controlling for algebra mathematics achievement. Hypothesis 4_b stated there would be no significant difference in the main effect of type of assessment on geometry achievement when controlling for algebra mathematics achievement. Hypothesis 4_c stated there would be no significant difference in the main effect of socioeconomic on geometry achievement when controlling for algebra mathematics achievement. The covariate of algebra I was dropped and a two-way factorial ANOVA was used to test the hypothesis. There was not a significant interaction between the independent variables SES and assessment type on the dependent variable mathematics achievement measured by the geometry end of course exam. SES and assessment type did not work together as a factor to influence geometry achievement. For assessment type, a significant difference in geometry achievement was found between students in the Non-TLI assessment group and those in the TLI assessment group. A significant difference was also found in geometry achievement for students based on SES favoring students the non-economic disadvantaged students.

As with previous findings in this study, the assessment type differences favored students at schools were the locally developed (traditional) interim assessment methods were used. As noted previously, Shepard (2009) referenced active engagement by teachers in the formulation of curriculum and assessment led to improved student achievement. Again, this could be a similar reason for this difference with geometry achievement. In addition to teacher buy-in, Brown and Hirschfeld's (2007) study linked student perception of teacher buy-in to student achievement, which points to the possibility that commercial programs may inhibit teacher buy-in.

Previous studies identify mathematics achievement difference based on SES favoring non-economically disadvantaged students (Myers, 1986). According to Johnson et al. (2006), economically disadvantaged students and parents may place less emphasis on mathematics education, and formal education as a whole. This typical socioeconomic difference for mathematics achievement in middle and high school can be overcome by participation in supplemental education services. (Heinrich et al., 2009). In this study, non-economic disadvantaged students had higher mean scores than economic disadvantaged students, which is in line with the findings of previous studies in this area.

Summary

Overall, in this study, the non-TLI students scored better than their TLI counterparts did in both the algebra and the geometry aspects of mathematics, even after the groups were leveled for previous mathematics achievement. Also, students who were non-economically disadvantaged performed better than those who were economically disadvantaged in geometry, but not in algebra (after controlling for previous mathematics achievement). Finally, in regards to gender, females were found to significantly outperform their male counterparts in geometry, when previous mathematics performance was accounted for, but not in algebra.

As mentioned earlier, the findings favoring the non-commercial (traditional) assessment programs are at the very least interesting when compared to previous findings. It is therefore worth mentioning that the TLI schools from which participants were drawn for this study were in an early stage of implementing the program. There is evidence in the literature to suggest that this type of assessment takes time to show significant improvement (Brown & Hirschfeld, 2007; Shepard, 2009). It is, therefore, plausible that the difference found in this study could just be an implementation dip as students at these schools adjust to the implementation of a new program. The case for an implementation dip is supported when considering the trend at these schools over the course of the two years before and after implementation. The three schools that began using TLI in the 2009-2010 (the year of data collection) school year had an overall drop in proficient or advanced scores of approximately 4% from the previous two years on the end of course algebra I exam and approximately 2% in geometry. For the two years after implementation of TLI, the three schools showed an average increase of approximately 1% in algebra I and approximately 6% increase in geometry. The positive effects of using interim assessment are realized when teacher have a thorough understanding of assessment according to Perie et al. (2009). In the first year of TLI, teachers use initial assessment module training that is two days to two weeks in length before the start of school (TLI, 2010).

The results of this study support the existence of a difference in mathematics achievement between the students who received instruction in a TLI based classroom and those who did not in six schools in the Arkansas River Valley. The six schools all used locally developed assessments in the previous school year. This difference remained even when the students algebra I scores were adjusted for their previous years eighth grade benchmark mathematics scores. Similarly, the results of the study confirm the existence of a comparable difference in geometry achievement between the students in a TLI classroom and those that did not. This difference remained even when the students geometry scores were adjusted for their previous years algebra I scores. There was also a difference in mathematics achievement observed between genders, however this was not consistent in the comparisons of algebra I and geometry scores. This difference was such that the difference seemed to show females scoring higher than males. One possible explanation for the existence of the difference in this case may be the higher mathematics achievement of female students at these grade levels, but this contradicts what has been observed by other researchers (Heinrich et al., 2009; Rutherford et al., 2010; Wilson & Zhang, 1998).

Furthermore, the results of this study show that students of different SES did not have a significant achievement difference in mathematics when comparing algebra I test scores while accounting for the previous year eighth grade benchmark scores. This appears to contradict studies by previous researchers (Heinrich et al., 2009; Johnson et al., 2006; Myers, 1986). There was an achievement difference observed in mathematics achievement between SESes when comparing geometry scores but not while adjusting for algebra I scores. The covariate of algebra I was removed for this comparison. This observation supported what was observed by other research (Heinrich et al., 2009; Johnson et al., 2006; Myers, 1986).

The results of this study indicated that gender and SES did not make a difference in algebra mathematics achievement, but did make a difference for geometry achievement during the first year of implementation of TLI assessment at three school districts in the Arkansas River Valley. These results do not fully correspond with the findings in the review of literature (Black & Wiliam, 2009, 2010; Cech, 2008). In this study, there was not a difference in mathematics achievement levels for algebra I for males and females although females showed a slight increase in achievement. However, there was a difference in mathematics achievement levels for geometry for males and females although females had a slight increase in mathematics achievement.

Recommendations

Therefore, the first recommendation is that school administrators should be cautious when starting a new assessment program. The use of TLI in this study did not appear to yield positive results in the first year of implementation, but school administrators may choose to see what happens in subsequent years of TLI. Although it appears that this program has negative effect on the students, the trend for two years after implementation show that scores are beginning to move in a positive direction. One could also argue that this small growth does not warrant such continued implementation. The effects of using TLI assessment within an established program should be studied for the long- range effects on mathematics achievement.

In any first year of implementation, teachers will experience a high learning curve, as they work to meet the learning needs of their students based on a new assessment. Both the teachers and the students had to adjust to the transition from Non-TLI assessment to TLI assessments. During the second year, teachers should be able to build upon their prior knowledge and ultimately be more successful meeting student needs and increasing mathematics achievement. Subsequent years of implementation with students continuing in TLI assessment classrooms may better represent the potential benefits of the program. In comparison with prior years, it was observed that overall mathematics achievement was lower for algebra and geometry in the first year of implementation of TLI assessment.

To fully understand the effects of the program, it may important not only to continue to offer TLI assessments, but to also extend the program to the other areas such as science and literacy. This could make the program more cohesive for students by having a systemic process in place for teachers and students in the major core areas. Students should be studied over time to see if TLI assessment utilization has a cumulative effect on student achievement. Does the time length of student participation in the assessment program strengthen the effect the assessment program has on student achievement? With some significant negative student achievement effects found in this study based on assessment type, it may prove valuable for schools to increase participation in this assessment program over a longer period and reassess.

A second recommendation is to continue professional development on understanding how to use assessment to drive instructional practices and build curriculum. The review of literature suggests that professional development is an important factor in the success of an effective assessment program (Ananda, 2003; Sheppard, 2009). There was limited formal professional development during this initial implementation year in the three schools. The schools had a short amount of time to build their modules and gain training on how to use the assessment and related reporting tools. There should be continued professional development to provide teachers with more tools to be successful in teaching students of both genders and SESes. At the beginning of this assessment program, teachers attended one to two day of training with a Learning Institute trainer. It is suggested that teachers continue the training started with TLI.

Significant difference seen in TLI and Non-TLI students in this research project do not follow the stereotypical ideas and national data discussed in the review in regards to gender and SES (Black & Wiliam, 2009, 2010; Mitchell, 1999; Myers, 1986; Scherer, 2010; Wilson & Zhang, 1999). The literature (Wilson & Zhang, 1998) indicated that males do better in mathematics achievement than females. However, the study indicated that this trend was not seen in all cases. Females seemed to do better than males in geometry mathematics achievement, but not a significant difference for algebra. A significant gender difference might indicate a need to extend professional development on understanding gender differences for teachers if this proved to be a significant difference in future studies with more longitudinal data.

Teachers of both genders could benefit from a better understanding of gender tendencies. If one's goal were to make all students successful, a better understanding of the learning differences between the male and females would give teachers more understanding on how to personalize learning plans for students and increase mathematics achievement.

In addition, more research should be done to understand how the differences in SES affect mathematics achievement. This study only looked at the student achievement aspect of the classroom based on mathematics achievement. A third recommendation is that future studies focus on mathematics achievement based on SES. Teachers could gain valuable insight from future studies and professional development that will help them in giving students quality instruction regardless of SES and concentration. Finally, teacher quality and student conceptions with assessment should also be studied.

Implications

Significance and Expansion of Knowledge Base

The first implication of this study is that significant differences in achievement based on assessment type may be evident during the first year of implementing the TLI program. This can be contributed to a change in assessments and result in this implementation dip. This was observed by looking at the prior algebra and geometry scores of the three schools that implemented TLI assessment for the first time. Although this study suggests that during the first year of implementing TLI a significant dip did occur, it will be important to check these results with other districts initiating similar programs. However, these findings should not be a deterrent to school districts that want to begin such a program. Over time there could be positive changes occur. Another implication of this study is that significant differences may not be evident on assessment when factoring in gender or SES. This could be contributed to sample size or the first year of implementation. Future yields could lead to significant differences due to gender or SES.

This study had several strengths. One strength was that it used closely matched schools as participants in the study. Another strength of this research was that at each of the TLI schools started the program at the same time. Students were taught the same standards, located in the same region, and exposed to the same type of module based instruction and assessment. Another strength is that by using this type of covariate based analysis, the adjusted means make the participants more comparable by adjusting for variation between students. The results of this research add to the growing body of research on assessment type within school districts in a high school setting. It also adds to the current studies on gender and SES effects on mathematics achievement. This study can serve as a starting place for future studies in the areas of assessment, gender, and SES.

Future Research Considerations

Although the focus of this study was student mathematics achievement, it is important to look at all aspects of the classroom when assessing the benefits of a program. A qualitative study that looks at the items of teacher satisfaction, student self-confidence, parent satisfaction, discipline, and graduation rates could prove to be helpful in showing the benefits of this type of assessment program. Various research methods could provide a comprehensive view of the overall effectiveness of the assessment program. This type of study could provide more insight to how the program is working and lead to future assessment strategies.

Additional research projects should be conducted on effects of assessment programs in high schools. There is not substantial research available that looks at the effectiveness of assessment programs in regards to high school mathematics achievement. Not enough is known about the effects of interim assessment programs within a public school setting, specifically at the high school. There is also not substantial research on how gender and SES influences assessment effectiveness. Wiliam and Black (2009) stated that if educators understood timeliness of feedback better, fewer students would fall behind in their learning. A better understanding timely feedback through quality assessment will lead to better mathematics achievement by students. Focusing feedback in a positive and timely manner can lead to students being more engaged in the educational process. Future studies could look at this correlation between timeliness of feedback and student satisfaction with their learning.

Myers (1986) related that some research studies indicate that low socioeconomic students can benefit most one on one teaching and timely feedback. More studies need to focus on the effects of SES on student mathematics achievement in both low and high concentration of poverty schools.

Potential Policy Changes

Ananda (2003) stated that policymakers, at the national level, ushered in the onslaught of testing within the public school setting through NCLB. These changes were made to assess the effectives of instruction through the measurement of mathematics achievement. Arkansas has developed an educational assessment program called ACTAAP within their state department of education to measure this achievement in mathematics, science, and literacy (Arkansas Department of Education, 2010). This assessment as has led schools to introduce many measures of assessments to prepares students for ACTAAP. This collection of assessment provides data to assess the effectiveness of these schools in their teaching and learning. States that do not currently take this approach should consider this direction for assessment.

Recently, states have undergone scrutiny for the number of assessments that schools are giving to students. There are always time costs when trying to measure the effectiveness of teaching and learning. States would be well served in taking an assessment inventory to measure the effectiveness of each type of assessment. This information could lead states in giving the most effective types of assessments that truly lead to better achievement by better instruction. Another key change might be to support schools in making assessment more organic to the typical learning day. This can be done through careful planning and embedding of assessment that is both time efficient and gives timely feedback to students and teachers. States could provide resources to develop assessments along with schools and reward schools with successful innovative assessment programs. Professional development can be set up for teachers to share new ideas and highlight effective ways of assessment that lead to effective teaching and learning. This should be the goal for all schools.

REFERENCES

Ananda, S. (2003). *Rethinking issues of alignment under "No Child Left Behind." knowledge brief*. Retrieved from ERIC database. (ED476416)

Arkansas Department of Education. (2010). *Student assessment.* Retrieved from http://arkansased.org/testing/assessment.html

Ayala, C., Shavelson, R., Ruiz-Primo, M., Brandon, P., Yin, Y., Furtak, E. … Ruiz-Primo (2008). From formal embedded assessments to reflective lessons: The development of formative assessment studies. *Applied Measurement in Education*, *21*(4), 315-334.

Bellamy, J. S., & Mativo, J. M. (2010). A different angle for teaching mathematics. *Technology Teacher, 69*(7), 26-28.

Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability*, *21*(1), 5-31.

Black, P., & Wiliam, D. (2010). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappan*, *91*(9), 81-90.

Brown, G., & Hirschfield, G. (2007). Students' conceptions of assessment and mathematics: Self-regulation raises achievement. *Australian Journal of Educational & Developmental Psychology*, *7*, 63-74.

Cech, S. (2008). Test industry split over "formative" assessment. *Education Week, 28*(4), 1.

Chappuis, S., Chappuis, J., & Stiggins, R. (2009). The quest for quality. *Educational Leadership*, *67*(3), 14-19.

Cobb, C., & Rallis, S. (2008). District responses to NCLB: Where is the justice? *Leadership and Policy in Schools, 7*(2), 178-201.

Davis, G., & McGowen, M. (2007). Formative feedback and the mindful teaching of mathematics. *Australian Senior Mathematics Journal*, *21*(1), 19-29.

Dunn, K., & Mulvenon, S. (2009). *Let's talk formative assessment and evaluation*. Retrieved from ERIC database. (ED 505357)

Gasbarra, P., & Johnson, J. (2008). *Out before the game begins: Hispanic leaders talk about what's needed to bring more Hispanic youngsters into science, technology and mathematics professions*. Retrieved from http://www.nassgap.org/library/docs/outbefore.PDF

Georges, A., Borman, K., & Lee, R. (2010). Mathematics reform and teacher quality in elementary grades: Assessments, teacher licensure, and certification. *Education Policy Analysis Archives*, *18*(13), 1-36.

Ginsburg, H. (2009). The challenge of formative assessment in mathematics education: Children's minds, teachers' minds. *Human Development*, *52*(2), 109-128.

Helen, C. (2011). Mathematics in the age of technology: There is a place for technology in the mathematics classroom. *Journal of The Research Center for Educational Technology*, *54* (1).

Heinrich, C., Meyer, R., Whitten, G., & Urban, I. (2009). *Supplemental education services under No Child Left Behind: Who signs up, and what do they gain?* Working Paper. Urban Institute, Retrieved from ERIC database. (ED508259)

Johnson, J., Arumi, A., Ott, A., & Remaley, M., (2006). *Are parents and students ready for more mathematics and science?* Retrieved from http://www.publicagenda.org/files/rc0601.pdf

Keegan, L., Orr, B. J., & Jones, B. J. (2002). Adequate yearly progress: Results, not process. In Thomas B. Fordham Foundation (Ed.), *No Child Left Behind: What will it take?* Retrieved from http://edex.s3-us-west-2.amazonaws.com/publication/pdfs/NCLBreport_7.pdf

The Learning Institute. (2010). *About the Learning Institute*. Retrieved from http://tlionline.net

Lee, V. E., Robinson, S. R., & Sebastian, J. (2012). The quality of instruction in urban high schools: Comparing mathematics and science to English and social studies classes in Chicago. *High School Journal*, *95*(3), 14-48.

Lesisko, L., Wright, R., & O'Hern, B. (2010). *Technology integration: A best practice perspective.* Retrieved from ERIC database. (ED508318)

Lord, H. G. (1973). *Ex post facto studies as a research method.* Retrieved from ERIC database. (ED090962)

Mathis, W. J. (2004). NCLB and high-stakes accountability: A cure or a symptom of the disease? NCLB: Failed schools--or failed law? *Educational Horizons*, *82*(2), 143-152.

Mertler, C. A., & Vannatta, R. A. (2013). *Advanced and multivariate statistical methods: Practical application and interpretation* (5th ed.). Glendale, CA: Pyrczak Publishing.

Mitchell, F. (1999, April). *All students can learn: Effects of curriculum alignment on the mathematics achievement of third-grade students.* Paper presented at the annual meeting of the American Educational Research Association, Quebec, Canada.

Myers, D. (1986). *The relationship between school poverty concentration and students' reading and mathematics achievement and learning.* Retrieved from ERIC database. (ED293955)

National Board Resource Center at Stanford University. (2010). *A quality teacher in every classroom: Creating a teacher evaluation system that works for California. Perspectives from accomplished California teachers.* Retrieved from http://nbrc.stanford.edu

Osta, I. (2007). Developing and piloting a framework for studying the alignment of mathematics examinations with the curriculum: The case of Lebanon. *Educational Research and Evaluation*, *13*(2), 171-198.

No Child Left Behind Act of 2001. (2002). Pub. L. No. 107-110, § 115, Stat. 1425.

Pearson Education. (2009). *Arkansas Comprehensive Testing, Assessment, & Accountability Program: Arkansas Augmented Benchmark Examinations: Mathematics, Literacy & Science. Technical report.* Upper Saddle River, NJ: Author.

Pearson Education. (2010). *Arkansas Comprehensive Testing, Assessment, & Accountability Program: Arkansas Augmented Benchmark Examinations: Mathematics, Literacy & Science. Technical report.* Upper Saddle River, NJ: Author.

Perie, M., Marion, S., & Gong, B. (2009). Moving toward a comprehensive assessment system: A framework for considering interim assessments. *Educational Measurement: Issues and Practice, 28*(3), 5-13.

Phillips, D. A., & Meloy, M. E. (2012). High-quality school-based pre-K can boost early learning for children with special needs. *Exceptional Children*, *78*(4), 471-490.

Rutherford, T., Kibrick, M., Burchinal, M., Richland, L., Conley, A., Osborne, K., … Martinez, M. (2010, May). *Spatial temporal mathematics at scale: An innovative and fully developed paradigm to boost mathematics achievement among all learners.* Paper presented at the annual meeting of the American Educational Research Association, Denver, CO.

Sawchuk, S. (2008). Leadership gap seen in post-NCLB changes in U.S. teachers. *Education Week*, *28*(3), 1.

Scherer, M. (2010). Getting a grip on the gaps. *Educational Leadership*, *68*(3), 7.

Seed, A. H. (2008). Redirecting the teaching profession in the wake of a nation at risk and NCLB. *Phi Delta Kappan*, *89*(8), 586-589.

Shepard, L. (2009). Commentary: Evaluating the validity of formative and interim assessment. *Educational Measurement: Issues and Practice, 28*(3), 32-37.

Strasser, N. (2010). Who wants to pass mathematics? Using clickers in calculus. *Journal of College Teaching & Learning*, *7*(3), 49-52.

Strobel, K., & Borsato, G. (2012). *Caring and motivating middle school classrooms.* Retrieved from http://gardnercenter.stanford.edu

Sunderman, G. L., Orfield, G., & Kim, J. S. (2006). Flawed assumptions: How No Child Left Behind fails principals. *Principal Leadership*, *6*(8), 16-19.

White, T., Loker, T., March, A., & Sockslager, K. (2009). Is NCLB closing the minority-majority achievement gap? *Communique, 37*(8), 1.

Wilson, L., & Zhang, L. (1998). *A cognitive analysis of gender differences on constructed-response and multiple-choice assessments in mathematics*. Retrieved from ERIC database. (ED420721)

Appendix

HARDING
UNIVERSITY
INSTITUTIONAL REVIEW BOARD

Status of Request for Exemption from IRB Review

(For Board Use Only)

Date: July 18, 2011

Proposal Number: 2011-60

Title of Project: The Effect of Assessment Method on End-of-Course Geometry and Algebra Achievement

Principal Investigator(s): Miguel Hernandez mhernandez@harding.edu

- [x] Research exempted from IRB review.
- [ ] Research requires IRB review.
- [ ] More information is needed before a determination can be made. (See attachment.)

I have reviewed the proposal referenced above and have rendered the decision noted above.
This study has been found to fall under the following exemption(s):

1 [ ] 2 [ ] 3 [ ] 4 [x] 5 [ ] 6 [ ]

In the event that, after this exemption is granted, this research proposal is changed, it may require a review by the full IRB. In such case, a ***Request for Amendment to Approved Research*** form must be completed and submitted.

This exemption is granted for one year from the date of this letter. Renewals will need to be reviewed and granted before expiration.

The IRB reserves the right to observe, review and evaluate this study and its procedures during the course of the study.

Chair, Harding University Institutional Review Board