The Effects of Curriculum-Based Measures on

Elementary Math Achievement

by

George Bickert

A Dissertation Presented in Partial Fulfillment
of the Requirements for the Degree
Doctor of Education

Approved March 2011 by the
Graduate Supervisory Committee:

Jere T. Humphreys, Chair
Nicholas Appleton
Dee Ann Spencer

ARIZONA STATE UNIVERSITY

May 2011

ABSTRACT

The No Child Left Behind Act of 2001 has had significant ramifications across public education. Due to reporting mandates, schools and districts are being held publicly accountable for the academic performance and progress of all students. Since implementation of the law, much attention has been focused on the “achievement gap,” that is, any differences in performance between groups of students. Students associated with the achievement gap typically come from certain demographics: minorities, low-income families, students with disabilities, and students with limited English proficiency (English Language Learners).

The purpose of this study was to examine the effect of using curriculum-based measures (CBMs) on math achievement, particularly ELL students. Eight elementary schools in northwestern New Mexico, divided into two groups (control and experimental) of four schools each, used the same state-approved, core math curriculum, were compared using a quasi-experimental research design. In addition to the regular core curricular materials, the experimental schools were provided with weekly CBMs, containing sample questions developed from the state’s performance standards. Each weekly CBM included at least one question from each of the five broad math strands: number and operations, algebra, geometry, measurement, and data and probability. Fourth (N = 283) and fifth grade (N = 294) students who had continuous enrollment for the duration of the experiment served as subjects. Successive regular administrations of the *New Mexico Standards Based Assessment* math subtest served as the pre- and posttest measures.

Analysis of covariance tests, with the pretest as the covariate, revealed no significant treatment effects for either the fourth or fifth grade students through the use of CBMs as a supplement to the core math curriculum. The significant effects, supported by previous research, were the school and, especially, the teacher for both grades. In this study, the effects of the classroom teacher were of more importance to student achievement than either the school a child attended or what curriculum program or process a given school employed.

To my family, colleagues, and students.

ACKNOWLEDGEMENTS

I would like to especially thank my mentor and advisor, Dr. Jere Humphreys, for guidance and encouragement throughout the process. I would also like to thank Dr. Nicholas Appleton and Dr. Dee Spencer, my committee members, for their time and support. I would like to give special acknowledgement to Dr. L. Dean Webb for making the Native American Educational Leadership (NAEL) Cohort II a reality.

I would like to thank (soon-to-be) Dr. Jacqueline Benally for the academic partnership throughout this project and overall friendship and support through the program. I would also like to acknowledge my other fellow NAEL Cohort members. I wish you all the best through your own dissertation projects and into the future.

I would like to especially thank Corazon Barsana, Dr. Jane Fernandez, Melody Pierson, and Elizabeth Thompson for their work and dedication on the original “Math Monsters” at Tohatchi Elementary School.

I would like to thank the administrators who believed in me, but more importantly, believed in trying to do something positive for students.

Finally, I would like to thank my wife, Janice, who sacrificed and supported me without complaint, through the entire process.

TABLE OF CONTENTS

Page

CHAPTER Page

LIST OF TABLES

LIST OF FIGURES

Figure Page

GLOSSARY OF ACRONYMS

AMO	Annual Measurable Objective
ANCOVA	analysis of covariance
AYP	Adequate Yearly Progress
BICS	Basic Interpersonal Communication Skill
CA1	Corrective Action 1
CA2	Corrective Action 2
CALP	Cognitive Academic Language Proficiency
CLD	culturally and linguistically diverse
CBM	curriculum-based measures
ELD	English Language Development
ELL	English Language Learners
ESEA	Elementary and Secondary Education Act
ESL	English as a second language
FAY	Full Academic Year
LBPA	language-based performance assessments
LEP	limited English proficiency
MC	multiple choice
NAEP	National Assessment of Educational Progress
NCLB	No Child Left Behind
NMELPA	New Mexico English Proficiency Assessment
NMPED	New Mexico Department of Education
NMSBA	New Mexico Standards Based Assessment
OE	open-ended
R-1	Restructuring-1
R-2	Restructuring 2
SA	short answer
SDAIE	Specially Designed Academic Instruction in English
SINOI	School In Need Of Improvement
SIOP	Sheltered Instruction Observation Protocol
SPSS	Statistical Package for the Social Sciences
WiDA	World-Class Instructional Design and Assessment

CHAPTER 1

INTRODUCTION

When English is the second language, immigrant and Native American students are expected to learn it as English is the official and dominant language in the United States. Education for both groups, immigrants and Native Americans, is used, in part, as a means of assimilation (Webb, 2006). These children continue to be expected to jettison their old ways and language and become English-speaking "Americans." Because some other languages, among them indigenous American languages, have different orthographic and lingual bases, second language learning and mastery are extremely difficult for students.

Since the beginning of the American experiment with publicly supported schools, there have been calls for improvement. By the early 1990s, President George H. W. Bush, along with the National Governors Association, agreed that the states should focus on improving academic standards and student achievement, and schools should be held accountable for the results (Webb, 2006). These recommendations were included in Bush's educational plan, *America 2000: An Education Strategy*, which was presented to but not enacted by Congress. Following the Bush administration, President Bill Clinton's administration crafted a plan of its own called *Goals 2000, Educate America Act*, which Congress did pass. This plan focused on issues like school readiness, student achievement, teacher education, mathematics and science, and lifelong learning.

Another reform, the No Child Left Behind (NCLB) Act of 2001, began following the election of President George W. Bush in 2000. NCLB was a reauthorization of the Elementary and Secondary Education Act (ESEA) of 1965. Taylor, Stecher, O'Day, Naftel, and LeFloch (2010) succinctly summarized the mission and focus of NCLB as all children will be proficient in reading and mathematics by the 2013-14 school year. They also described further requirements of the Act, stating that schools and districts will be held accountable for their students' progress and mastery of state academic standards, as measured by state tests. Further, students with limited English proficiency (LEP) and students receiving special education services are to be included and reported separately. The five specific areas addressed in NCLB are (a) proficiency standards in reading, writing, and mathematics; (b) standards and requirements for highly qualified teachers; (c) proficiency standards for Limited English Proficiency (LEP) students; (d) requirements for safe and drug-free schools; and (e) high school graduation rate requirements. Since NCLB became law, the accountability reporting data have shown a clear achievement gap between LEP and non-LEP students.

Across the country, and within states' NCLB accountability reports, the LEP label is being replaced with the term *English Language Learners* (ELL). As used increasingly in New Mexico and across the country, the ELL designation applies to an individual who is in the process of acquiring skills in the English language and whose first language is not English. It is a new singular title given to several unique groups. Other terms commonly found in the literature include *language minority students*, *limited English proficient*, *English as a second language* (ESL), and *culturally and linguistically diverse* (CLD). Bank Street College (n.d.) identified ELL as the new label for students whose second language is English. This shift in labeling represents a more accurate reflection of the process of language acquisition.

The federal law focuses on promoting English language development and providing appropriate grade-level academic content to students. NCLB includes requirements that states establish standards and benchmarks for English language proficiency and academic content. According to Webb (2006), with NCLB, the most sweeping education reform legislation since the ESEA, the Bush administration did what a Democratic administration could never have done: created "a much larger federal presence in educational policy and funding and set the foundation for a national testing system. NCLB provided the framework and impetus for standards-based reform of education in state after state" (Lewis, quoted in Webb, 2006, p. 184). In addition, NCLB designated "English Language Learners" as a demographic subpopulation that must be assessed and meet Adequate Yearly Progress (AYP) goals. State departments of education are required to complete an AYP analysis for all public schools and districts that serve such students.

The foreign-born population of the United States has tripled over the past 30 years, with more than 14 million immigrants entering the country during the 1990s alone. Passel and Cohn (2008) projected that up to 150 million additional immigrants will be added to the population between the years 2005 and 2050. If this happens, the ELL population will increase dramatically as well.

Currently, the ELL population is the fastest growing segment of the student population, with the highest level of growth occurring in grades 7–12, where ELL students increased by approximately 70% between 1992 and 2002. ELL students now comprise 10.5% of the nation's K–12 enrollment, up from 5% in 1990. ELL students do not fit easily into simple categories because they comprise a highly diverse group. Recent research shows that 57% of adolescent ELL children were born in the U.S., while 43% were born elsewhere. ELL students have various levels of language proficiency, socio-economic standing, schooling, content knowledge, and immigration status (National Council of Teachers of English, 2008).

While the term ELL might suggest a homogeneous group, ELL children are a highly heterogeneous and complex group of students with diverse gifts, educational needs and backgrounds, languages, and goals. Some ELL students come from homes in which no English is spoken, while others come from homes where only English is spoken. Still others have been exposed to or use multiple languages. ELL students may have a deep sense of their non-U.S. culture, but most have a strong sense of multiple cultures and identities. Some ELL students are stigmatized for the way they speak English, while some are stigmatized for speaking a language other than English (National Council of Teachers of English, 2008).

Many ELL students go through a "silent period" during which they listen and observe more than they speak. During this silent period, ELL students benefit from opportunities to participate and interact with others in activities that use gesture(s), physical movement, art, experiential activities, and single words or short phrases. Most ELL students acquire the ability to understand and use the predictable oral language needed for daily routines, play, and social interaction before they develop the ability to understand and use oral and written academic English. Unfortunately, this discrepancy between Basic Interpersonal Communication Skill (BICS) and Cognitive Academic Language Proficiency (CALP) is not widely understood (Schon, Shaftel, & Markham, 2008).

BICS describes the language skills needed in social situations. ELL students employ BICS when they are on the playground, in the lunch room, on the school bus, at parties, playing sports, and talking on the telephone. Social interactions are usually context-embedded, which means that they occur in a meaningful social context. These language skills are not very demanding cognitively, do not require specialization, and usually develop within six months to two years after arrival in the U.S.

CALP refers to the formal academic language needed for classroom learning. These skills include listening, speaking, reading, and writing about subject area content material. This level of language learning is essential for student success in school. Students need time and support to become proficient in academic areas, a process that typically takes from five to seven years. Thomas and Collier (2002) reported that if a child has no prior schooling or other types of support in native language development, it may take seven to ten years for ELLs to catch up to their peers.

Academic language acquisition includes more than just understanding content area vocabulary; it also includes skills such as comparing, classifying, synthesizing, evaluating, and inferring. Academic language tasks are context-reduced in comparison to social language, as information is read from a textbook or presented by the teacher. In addition, the language becomes more cognitively demanding as new ideas, concepts, and vocabulary are presented simultaneously to students. Cummins (1984) developed a useful BICS/CALP matrix for educators that represents a continuum from “context-embedded” to “context-reduced.” He also advanced a theory that there is a common underlying proficiency (CUP) between two languages. Skills, ideas, and concepts students learn in their first language will be transferred to the second language. However, specifics about how the process of the transference occurs are still unknown (Cummins, 1984).

When students who seem to speak and understand English well in daily life fail to perform well academically, a frequent assumption is that they have special needs or lack motivation. In fact, many ELL students are simply at a developmental stage in which they have acquired interpersonal language but cannot yet fully understand or express more complex thoughts in English. These students need numerous opportunities to listen, speak, read, and write across the curriculum. With time and adequate opportunities to listen, observe, participate, and interact, ELL students are able to progress in understanding and produce language that is increasingly understandable, complete, and grammatical (Cummins, 1984).

Each state has some leeway in how it measures and reports ELL students' academic achievement. The state of New Mexico measures each student's (grades 3-8 and 11) academic proficiency in reading and mathematics annually on an assessment called the *New Mexico Standards Based Assessment* (NMSBA). In addition to measuring core academic proficiency in reading and math, the state is also required to report on ELL students' progress and proficiency in English language skills.

Prior to the 2009-10 school year, New Mexico used the *New Mexico English Language Proficiency Assessment* (NMELPA). Currently, New Mexico has joined the World-Class Instructional Design and Assessment (WiDA) Consortium for the specific purpose of aligning its ELL proficiency assessments and procedures with those of other states. Currently, WiDA members include Alabama, Delaware, the District of Columbia, Georgia, Illinois, Kentucky, Maine, Mississippi, Missouri, New Hampshire, New Jersey, New Mexico, North Carolina, North Dakota, Oklahoma, Pennsylvania, Rhode Island, South Dakota, Vermont, Virginia, Wisconsin, and Wyoming. The initial and annual English proficiency progress monitoring assessment these states administer is called ACCESS for ELLs, or simply ACCESS. The assessment measures four areas or components of English proficiency: speaking, listening, reading, and writing.

Returning to NCLB and AYP, states must disaggregate school and district student scores to show proficiency in each subgroup (including ELLs). The New

Mexico Department of Education has developed Annual Measureable Objectives (AMOs) to ensure that the percentages of students scoring proficient (or advanced) on the NMSBA are sufficient for a school to make AYP and progress toward the ultimate goal of 100% proficiency by 2013-14. AMOs differ by subject and grade, but not by subgroups. Even before measuring performance against the established AMOs, schools must assess a minimum of 95% of Full Academic Year (FAY) students. Schools must also achieve an attendance rate of greater than 92%.

Although NCLB has focused on equalizing educational opportunities for poor, minority, and at-risk children and the intent is to leave no child behind, the unfortunate reality is that many students are being left behind, especially students the law was intended to help. The law contains provisions that permit states to direct and focus more attention on low-achieving students and to intensify efforts to improve consistently low-performing schools. Peregoy and Boyle (2005) stated that the current emphasis on curriculum standards and high-stakes testing as required by NCLB has placed tremendous pressure on students, teachers, and administrators for ELL students to test well. Although NCLB targets poor and minority children, it also attempts to ensure that every child will be taught by a highly qualified teacher and will reach proficiency on a state-adopted achievement test. NCLB embodies test-driven accountability and has been a major influence on public schools nationwide.

Problems related to the achievement of ELL students are well documented. The U. S. Department of Education reported in 2009 that ELL students lagged significantly behind their peers and that the achievement gap was widening (Edwards, 2009). Also, Edwards reported a significant achievement gap at the national level between the ELL and the “all students” groups in *Education Week,* which devoted its entire annual “Quality Counts” issue to ELL issues. On the National Assessment of Educational Progress examination, only 9.6% of ELL students scored proficient in math, while 34.8% of the “all students” group was proficient. For some students, mathematics seems to be a foreign language, consisting of words and concepts that do not mesh with their everyday experiences. Mathematic classes for ELL students can be especially challenging because of the need to learn mathematics and English at the same time.

Edwards (2009) in *Education Week* also pointed out and described problems and challenges experienced by ELL students throughout the country: immigration, screening, professional development, funding, quality of instruction, and materials. The researchers proposed that the availability and use of quality standards-based weekly assessments are critical to the task of improving the academic achievement of ELL students, especially in mathematics.

Most research studies and essays on the impact of NCLB on public schools include the following themes: inadequate school funding to carry out the NCLB testing and accountability mandates; the challenge of meeting requirements to provide highly qualified teachers; difficulty in implementing research-based instructional practices; and attaining sufficient student achievement and proficiency levels in reading, writing, and mathematics. Measuring the effectiveness of a program designed to increase the percentage of

ELL students who are proficient in math on the NMSBA was the primary purpose of this research study.

Statement of the Problem

This study was designed to investigate the effects of using weekly curriculum-based measures (CBMs) on math achievement among elementary ELL students. Math achievement is defined herein as scores on the *New Mexico Standards Based Assessment*. Mathematics is a complex subject that ranges from simple addition to calculus. Because it was not feasible to cover the entire mathematics continuum, the study focused on ELL students enrolled in several New Mexico elementary schools.

Need for the Study

The achievement gap between ELL and non-ELL students necessitates the need to develop and then conduct research on programs and strategies that potentially could help close the gap. The results of this study have the potential to help New Mexico public schools close the achievement gap and realize the ultimate goal of NCLB. Effective weekly curriculum-based measures may become an additional “high-yield” strategy for working with ELL students. The result may also serve to encourage schools to revise their school improvement plans, curriculum, and testing procedures that affect not only ELL students, but also other students who fall into the achievement gap.

Delimitations

This study was limited to curriculum-based measures being used in math. Although NCLB has the potential to affect education in a variety of ways, this study was limited to the following: (a) assessment requirements on curriculum and instructional practices (math), and (b) requirements for meeting the needs of ELL students in the area of mathematics.

Research Questions

This study addressed the following research questions:

1. What is the impact of curriculum-based measures on the math achievement of elementary students?
2. What is the impact of curriculum-based measures on the math achievement of ELL students in particular?

Hypotheses

To answer the primary research questions, specific hypotheses were developed and tested. Four null hypotheses were developed for both subject groups. The following four null hypotheses were developed for the fourth grade subjects:

Null hypothesis 1a: There will be no statistically significant difference in math scores on the *New Mexico Standards Based Assessment* between the experimental and control groups ($p < .05$). This hypothesis was tested with an analysis of covariance test.

Null hypothesis 1b: There will be no statistically significant difference in math scores on the *New Mexico Standards Based Assessment* among English Language Learner groups ($p < .05$). This hypothesis was tested with an analysis of covariance test.

Null hypothesis 1c: There will be no statistically significant difference in math scores on the *New Mexico Standards Based Assessment* among schools ($p < .05$). This hypothesis was tested with an analysis of covariance test.

Null hypothesis 1d: There will be no statistically significant difference in math scores on the *New Mexico Standards Based Assessment* among individual teachers ($p < .05$). This hypothesis was tested with an analysis of covariance test.

The following four null hypotheses were for the fifth grade subjects:

Null hypothesis 2a: There will be no statistically significant difference in math scores on the *New Mexico Standards Based Assessment* between the experimental and control groups ($p < .05$). This hypothesis was tested with an analysis of covariance test.

Null hypothesis 2b: There will be no statistically significant difference in math scores on the *New Mexico Standards Based Assessment* among English Language Learner groups ($p < .05$). This hypothesis was tested with an analysis of covariance test.

Null hypothesis 2c: There will be no statistically significant difference in math scores on the *New Mexico Standards Based Assessment* among schools ($p < .05$). This hypothesis was tested with an analysis of covariance test.

Null hypothesis 2d: There will be no statistically significant difference in math scores on the *New Mexico Standards Based Assessment* among teachers ($p < .05$). This hypothesis was tested with an analysis of covariance test.

Definition of Terms

Adequate Yearly Progress (AYP): Under NCLB each state establishes a definition of AYP to use to determine the achievement of each school district and school on an annual basis. States are to identify for improvement any Title I school that does not meet the state's definition of adequate yearly progress for two consecutive school years.

Curriculum-based measures (CBM): Curriculum-based measures are assessments created from or aligned to the curriculum, used to measure student performance and progress on curriculum.

English Language Learner (ELL): The term *English language learner* as used in this study indicates a person who is in the process of acquiring English language skills and whose first language is not English. Other terms commonly found in the literature include *language minority students, limited English proficient* (LEP), *English as a second language* (ESL), and *culturally and linguistically diverse* (CLD).

New Mexico Standards Based Assessment (NMSBA): The official AYP test of the state of New Mexico. The test assesses the New Mexico Content Standards in reading, writing, math, and science for grades 3-8, and 11. In grade 11 the test also assesses social studies.

No Child Left Behind Act of 2001 (NCLB): The reauthorization of the Elementary and Secondary Education Act of 1965 that was the force of federal law affecting K-12 schools at the time of this study.

Organization of Chapters

Chapter 1 presents the problem of achievement gap between ELL and non-ELL students and explains that this study was designed to investigate the effects of using weekly curriculum-based measures on math achievement, particularly among elementary ELL students. Research presented in chapter 2 substantiates the continuing problems and the many and varied programs of developing curriculum for ELL students and concludes that there is no published research on the use of intervention programs or strategies to improve ELL students' mathematics achievement on language-based performance assessments. Presented in chapter 3 is the methodology used to assess the results of fourth and fifth grade students in eight public elementary schools from one school district in New Mexico using the *New Mexico Standards Based Assessment*. Two sections in chapter 4 present in narrative and table formats the results of using CBMs. The main finding that teachers had a significant effect on math achievement concludes Chapter 5.

CHAPTER 2

REVIEW OF THE LITERATURE

Defining English Language Learners

Limited English proficiency was defined in the Elementary and Secondary Education Act of 1965 as

> an individual, means an individual (A) who is aged 3 through 21; (B) who is enrolled or preparing to enroll in an elementary school or secondary school; (C)(i) who was not born in the United States or whose native language is a language other than English; (ii)(I) who is a Native American or Alaska Native, or a native resident of the outlying areas; and (II) who comes from an environment where a language other than English has had a significant impact on the individual s level of English language proficiency; or (iii) who is migratory, whose native language is a language other than English, and who comes from an environment where a language other than English is dominant; and (D) whose difficulties in speaking, reading, writing, or understanding the English language may be sufficient to deny the individual (i) the ability to meet the State's proficient level of achievement on State assessments described in section 1111(b)(3); (ii) the ability to successfully achieve in classrooms where the language of instruction is English; or (iii) the opportunity to participate fully in society. (Section 9101(25)).

The subsequent shift in labeling from LEP and ESL to ELL for students with limited English proficiency provides a more accurate reflection of the process of language acquisition because these students are in various stages of acquiring English skills. Peregoy and Boyle (2005) provided a more detailed description of students who may currently fall under the ELL umbrella. They say that ELL students may be the children of immigrants coming to the U.S. looking for a better economic life, looking to escape war or political unrest in their native countries, or even be children who were born here, such as Native Americans, whose "roots in American soil go back for countless generations" (p. 2). Regardless of when, why, or how these students arrived in American public schools, their commonality is that all speak a primary language other than English in the home but are required to learn and show their proficiency in academic areas, especially reading and math, through use of the English language.

History of English Language Learners

In 1974, the U.S. Supreme Court ruled in *Lau v. Nichols* that failure to provide appropriate educational support for students with limited English proficiency violates students’ rights. The need to provide and monitor services and educational progress of ELL students has been recognized continually by the courts and by legislation (i.e., No Child Left Behind Act of 2001).

The federal No Child Left Behind Act of 2001 not only requires schools, districts, and states to identify and track ELL students, it also mandates that ELL students be reported as a unique subpopulation for determination of adequate yearly progress. However, the NCLB Act does not specifically define what constitutes ELL. The process of identifying ELL students is left to the individual states.

Identification of English Language Learners

Goldenberg and Rutherford-Quach (2010) studied the identification of ELL students nationwide and found that while the process varies from state to state, a two-step process is typical. The first step involves an initial report, referral, or indication that a student might have limited English proficiency. Step two involves administering an English proficiency test to make an identification.

Many states use a “Home Language Survey” as the primary means of identifying potential cases of limited English proficiency. A few years ago

Kindler (2002) reported that 45 states used the survey as an identification tool for determining limited English proficiency. More recently, *Education Week* reported that 49 of the 51 states (including the District of Columbia) use a home language survey as part of the referral process (“Identifying English-language Learners,” 2009). Although home language surveys are often used as the first measure of a potential English proficiency problem, they are fraught with controversy. First, these surveys tend to be simple, usually just asking questions about languages spoken in the home and perhaps one or two other language-related questions. Abedi (2008) questioned the reliability and validity of these surveys, stating that there is no relationship between parents’ responses on the surveys and students’ measured proficiency levels. Second, Littlejohn (1998) argued that the use of these surveys over-identifies students in the ELL category because not all students who have a home language other than English have limited English proficiency.

After an initial referral of a potential English proficiency problem, all states and the District of Columbia determine ELL status by assessing the referred students on English proficiency. Again, there is no universal or national criteria for these assessments; instead, states have the right and obligation to create their own.

In 2009, New Mexico joined the World-Class Instructional Design and Assessment Consortium for the specific purpose of aligning the state’s ELL proficiency assessments and procedures with those of other states. These states administer initial and annual English proficiency assessments called ACCESS for ELLs, or simply ACCESS. The assessments cover four areas or components of

English proficiency: speaking, listening, reading, and writing. Prior to 2009, New Mexico developed and administered its own test called the New Mexico English Language Proficiency Assessment, which measured English proficiency in the areas of speaking, listening, reading, and writing. No official statement or comment has been provided by the New Mexico Public Education Department on why it chose to stop using its own test and join the WiDA consortium.

Curriculum, Instruction, Assessment

As part of NCLB's comprehensive accountability system, states are required to construct a system that measures schools' progress in reading and math (some states also include science). States have developed assessments in these areas aligned with their own state standards. Results on these assessments determine whether a school has made AYP. In New Mexico, a school making AYP is categorized as "Making Progress," while a school not making AYP is categorized as a "School In Need Of Improvement" (SINOI). There are four categories or ratings within the larger category of SINOI. The four ratings are (in descending order): Corrective Action 1 (CA1), Corrective Action 2 (CA2), Restructuring 1 (R-1), and Restructuring 2 (R-2). In addition to school ratings and designations, NCLB requires states and individual schools to monitor and track the progress of ELL students in reading and math.

Curriculum

Once individual students are classified as ELL, federal law requires that educational programs provide them with two components: access to the core curriculum and opportunities for English language development. Federal law makes no determination about how schools and districts will accomplish these two requirements. Instead, state education agencies and state laws govern program and curriculum implementation. Historically, since many LEP/ELL students were foreign-born, and in local areas spoke the same first language, services were provided under the aegis of *bilingual education.*

Lessow-Hurley (2000) discussed dual or bilingual education at length, concluding that all forms of bilingual education focus on teaching and improving English language skills, and on providing access to the core curriculum through the home language while learning English. Within bilingual education, Lessow-Hurley lists the most common types of programs: transitional bilingual, maintenance bilingual, immersion, two-way immersion, and newcomer programs.

Peregoy and Boyle (2005) reported that bilingual education programs serve only a small percentage of ELL students, and that the vast majority of students receive services through English Language Instructional Programs. This shift mirrors demographic changes in the population. Previously, ELL students in a given school or district tended to be from the same place and speak the same language. Today, schools and classrooms contain students from multiple locations who speak a multitude of languages. The four most common types of ELL instructional programs are Sheltered English, also referred to as Specially Designed Academic Instruction in English (SDAIE); ESL Pullout; English Language Development (ELD); and Structured English Immersion (Peregoy & Boyle, 2005).

At the time of this study, the frequency and type of programs ("Providing Language Instruction," 2009) offered by states are Content-based ESL (43 states, including the District of Columbia), Pull-out ESL (42), Sheltered English Instruction (39), Structured English Instruction (32), Dual language (31), Transitional bilingual (28), Two-way immersion (23), Specially designed academic instruction in English (18), Heritage language (16), Developmental bilingual (15), and Other (29).

In Sheltered English/SDAIE programs, subject matter is taught entirely in English, while the instructional approach, which includes specialized techniques, is designed to foster second language acquisition. With ESL Pullout programs, students receive the majority of their instruction in English but are "pulled out" of the regular class to receive help from an ESL teacher or assistant. ELD programs are very similar to SDAIE programs in that students receive all their instruction in English from teachers with special training in second language acquisition skills. Finally, in structured English Immersion, sheltering techniques are used to make the English-only content understandable.

In New Mexico, the Sheltered Instruction Observation Protocol developed by Echevarria, Vogt, and Short (2000) is the dominant ELL instructional approach. The authors described it as an approach to teaching that extends the time students have for receiving English language support while they learn content subjects. In this model, teachers build or scaffold instruction to increase comprehension of the material by adjusting their speech and instructional tasks, and by providing appropriate background information and experiences.

Instruction

The first requirement in terms of instruction for ELL students under NCLB is that they have access to the core curriculum. In general, all states have adopted a standards-based curriculum and focus instruction on standards within core subjects. Laturnau (2003) detailed the three components of standards-based instruction: (a) the content standards describe what students should know and be able to do; (b) benchmarks within the standards specify expected knowledge and skills for each standard at different grade levels; and (c) performance and progress indicators describe how students will demonstrate that they have met the standard.

The second requirement is that ELL students have opportunities to develop English language skills. The different types of opportunities currently in use were discussed above in the section titled "Curriculum." The Sheltered Instruction Observation Protocol (SIOP) proposed by Echevarria, Vogt, and Short (2000) and used as a curricular framework in New Mexico also contains and recommends particular instructional strategies. The strategic core of SIOP is for teachers to modify their language, making instructional talk more understandable by speaking clearly, repeating main ideas and key points, and defining needed vocabulary within context. Another important component is to combine verbal with nonverbal communications, such as gestures, graphs, pictures, and objects.

Peregoy and Boyle (2005) provided information on other high-yield ELL instructional strategies including group work, thematic instruction, and scaffolding. Many of these strategies are also included in the SIOP approach. In summary, for ELL students to achieve greater and deeper understanding and retention of material, instruction must combine comprehensible input with social interaction opportunities, such that ELL students can process information verbally and nonverbally. Gibbs (1994) also suggested that social interaction and positive relationships help promote success among ELL and other students.

Assessment

Assessment refers to any evaluation of student learning, progress, achievement, or development. For ELL students, two highly distinct assessment determinations occur every year, both mandated by NCLB. The first required assessment measures ELL students' progress in English language development. The second requires reporting on the progress of ELL students in the core curriculum subjects of reading and math (some states also require and publish information on progress in science). The primary goal of the first requirement is to have students reach proficiency in English, although general progress is also measured and reported. Every ELL student is assessed annually with the same instrument that originally indicated limited English proficiency for that individual. In New Mexico, ELL students are administered the ACCESS test. NCLB requires states to report annually on the English language development progress of ELL students.

Education Week ("Moving Toward Language Proficiency," 2009) reported the most up-to-date national information compiled by the U.S. Department of Education. Results indicated that there were almost 4.5 million students classified as ELL in 2006-07. Nationally, only 12.5% of those students tested as English proficient at the end of the year. In Arizona 10.7% of the ELL population reached proficiency levels (*N* = 167,679), while New Mexico advanced 22.5% of its ELL population (*N* = 62,812) to proficiency. The same report also provided information on the extent to which individual students progressed. Nationally, 34.4% of all ELL students made progress toward English proficiency. The results for Arizona and New Mexico were 47.8% and 30.0%, respectively ("Moving Toward Language Proficiency," 2009).

The second reporting requirement answers the question of how ELL students are faring in the core content areas of reading and math. Nationally, all students in grades 3–8 and 11 are assessed annually to measure their proficiency in reading and math. The results from this test determine a school's AYP status and thus are at the heart of NCLB's accountability requirements.

In New Mexico, all students in grades 3–8 and 11 take the *New Mexico Standards Based Assessment* (NMSBA), which measures performance in reading and math. Statewide performance data showing comparisons between ELL students and ALL students for 2008 and 2010 are displayed in Tables 1 and 2.

Achievement Gap

Christie (2002) provided a detailed definition of what is referred to in education as the *achievement gap.* It can be defined as a significant performance difference in an area (or areas) of a state test among any of various groups of students. Student groups can include male and female students, students with and without disabilities, students with and without proficiency in English, minority and nonminority students, and students who are eligible for free and reduced-price lunch versus those who are not eligible. For the purposes of reporting AYP, students without English proficiency are called ELL, and students eligible for free and reduced-price lunch are called Economically Disadvantaged.

Fry (2008) reported that analysis of assessment data uniformly indicates that ELL students are much more likely than non-ELL students to score below proficiency levels in both reading and math. McBride's (2008) report concurs with Fry's findings, and states further that ELL students constitute one of the lowest scoring groups on both national and state assessments. Additionally, she found that from 2005 through 2007 the achievement gap increased between non-ELL and ELL students on the National Assessment of Educational Progress examination.

Table 1 shows the achievement gap in reading between ALL students and ELL students on the NMSBA between 2008 and 2010. In all grade levels tested, there is at least a 10-point gap, a gap that has widened for every grade from 2008 to the present.

Table 1

New Mexico's Reading Achievement Gap

Grade	2008	2009	2010
3	16.2	15.7	17.6
4	16.7	18.2	18.7
5	21.1	23.7	23.6
6	17.8	20.1	19.8
7	19.6	24.4	24.8
8	18.7	27.3	28.3
11	25.0	27.8	30.1

Note. The values represent the difference between the percentage of ALL students and ELL students scoring proficient or above on the reading portion of the *New Mexico Standards Based Assessment*, in favor of ALL students in every case.

Table 2 shows the achievement gap in math between All and ELL students on the NMSBA between 2008 and 2010. For all grade levels tested, there is at least a 10-point gap, a gap that has widened for every grade from 2008 to the present.

Table 2

New Mexico's Math Achievement Gap

Grade	2008	2009	2010
3	7.3	7.1	10.2
4	13.2	14.2	17.2
5	15.7	17.7	19.8
6	13.0	16.5	17.5
7	16.9	18.2	20.8
8	19.3	23.0	24.3
11	21.7	22.0	24.8

Note. The values represent the difference between the percentage of ALL students and ELL students scoring proficient or above on the math portion of the *New Mexico Standards Based Assessment*, in favor of ALL students in every case.
Table 2

Non-Math Interventions

Numerous research studies have been conducted on programs, strategies, and interventions designed to improve the reading levels and overall academic achievement of ELL students. Within the domain of language arts, one can find research on improving comprehension (Basurto, 1999; Echevarria, Vogt, & Short, 2007; Haager & Windmueller, 2001; Kamps et al. 2007; Wright, 1997), vocabulary (Carlo, August, McLaughlin, Snow, Dressier, & Lippman, 2004), academic language (Bunch, Abram, Lotan, & Valdes, 2001; Dong, 2004), and literature (Ernst-Slavit, Moore, & Maloney, 2002; Gersten & Jimenez, 1994). General reading intervention studies focusing on ELLs have shown benefit when the instruction includes three factors in combination: (a) the essential components of reading, namely phonics, phonological awareness, fluency building, vocabulary, and comprehension; (b) the use of effective instructional practices; and (c) the development of English (Carlo et al., 2004; Denton, Anthony, Parker, & Hasbrouck, 2004; Gunn, Biglan, Smolkowski, & Ary, 2000; Vaughn et al., 2006). Shanahan and Beck (2006) concluded that early reading interventions used with non-ELL students also work with ELL students. However, Snow (2006) synthesized the scope of the research by the National Literacy Panel on Language-Minority Children and Youth by saying, "The literature reviewed reveals remarkably little about the effectiveness of different aspects of instruction, and provides only limited guidance about how good instruction for second-language speakers might differ from that for first-language speakers" (p. 638). He continued, "Most discouraging, the research we reviewed provides little basis for deciding whether or what kinds of accommodations or adaptations are most helpful to second-language learners" (p. 639).

Research in the field of science has focused on inquiry (Amaral, Garrison, & Klentschy, 2002; Cuevas, Lee, Hart, & Deaktor, 2005; Fradd & Lee, 1999), professional development for teachers (Buck, Mast, Ehlers, & Franklin, 2005; Hart & Lee, 2003), and lesson adaptations and accommodations (Rice, Pappamihiel, & Lake, 2004). In social studies, the focus has been on integration

(Short, 1994; Weisman & Hansen, 2007), and support (Brown, 2007; Egbert & Simich-Dudgeon, 2007; Szpara & Ahmad, 2007).

Math Interventions

One of the most important differences between math and the other core subjects (i.e., English, social studies, and science) is that math has its own unique language and symbols. Halliday (1978) was the first to coin the term *mathematics register*. He further defined registry as “a set of meanings that is appropriate to a particular function of language, together with the words and structures which express these meanings” (p. 195). Schleppegrell (2007) added that “learning the language of a new discipline is part of the learning of the discipline; in fact, the language and the learning cannot be separated” (p. 140). He stated further that there are three distinct linguistic challenges associated with math: (a) multi-semiotic formations, (b) dense noun phrases that participate in relational processes, and (c) precise meanings of conjunctions and implicit logical relationships that link mathematic elements. His conclusion is that “the linguistic challenges of math need to be addressed for students to be able to construct knowledge about math in ways that will ensure their success” (p. 156). Other researchers have simplified the language/semiotic debate and concluded that math language differs enough from everyday language that it presents challenges to all groups of students, and to ELL students in particular (Adams, 2003; Pimm, 1987; Spanos, Rhodes, Dale, & Crandall, 1988).

The importance of math language cannot be over-emphasized because, as Lager (2004) pointed out, “The more advanced math becomes the more language-dependent it is" (p. 1). Cardenas, Robledo, and Waggoner (1988) reported that the highest correlation with staying in school is enrollment in advanced math classes, while Wang and Goldschmidt (1999) reported that students who take elective math classes have the highest overall academic growth rates. Other researchers have consistently found a connection between the mathematics courses taken (or courses assigned) by students and their math achievement (Bryk, Lee, & Smith, 1990; Hoffer, Rasinski, & Moore, 1995; McDonnell, Burstein, Ormseth, Catterall, & Moody, 1990; Oakes, Gamoran, & Page, 1992; Rock, & Pollack, 1995; Wang & Goldschmidt, 2003).

Buchanan and Helman (1997) recommended that teachers not only teach the vocabulary of math, but explain the nuances of the language. (For example, when teaching *greater*, a teacher might also have to explain the meaning of the suffix *er*.) Lemke (2003) proposed that because semiotics creates such a problem for ELL students, teachers need to help them see natural language, mathematics, and visual representations as part of a unified system. Tevebaugh (1998) showed that ELL students would be more successful in math with extra math language instruction. Sfard, Nesher, Streefland, Cobb, and Mason (1998) recommended that teachers verbally explain the meanings of math symbols to facilitate better understanding, and suggested that focusing on the linguistic features could help clarify the technical meanings.

This recommendation of explaining meanings was supported by other research (O'Hallaran, 2000). Leung, Low, and Sweller (1997) found that until students gain experience and facility with solving problems, the teacher's verbal explanations are the most important component of instruction. Moschkovich (1999) concluded that to increase ELL students' language proficiency and achievement in math, students need to participate both orally and in writing by "explaining solution processes, describing conjectures, providing conclusions and presenting arguments" (p. 11). Other studies have clearly pointed to the significance of reading in terms of overall math performance and achievement (Helwig, Rozeck-Tedesco, Tindal, Heath, & Almond, 1999; Lager, 2006).

Two studies published in 1988 showed that the low math scores (achievement) of ELL students were a function of language, and that the scores could be improved by increasing students' language comprehension and by modifying the language of the assessment items (Cocking & Chipman, 1988; Mestre, 1988). Staub and Reusser (1995) supported these recommendations and showed that the wording of math problems has a major influence on comprehension and students' ability to solve problems. In a follow-up study, one of the most highly publicized and notable studies, Abedi and Lord (2001) modified the wording of math items on the National Assessment of Educational Progress and found that the linguistic modifications resulted in higher scores for ELL students. The argument continues to be made that high-stakes assessments are inappropriate for ELL students because of the way the tests are constructed and worded (Solorzano, 2008).

In a much less-publicized study, Abedi, Courtney, Leon, Kao, and Azzam (2006) found that ELL students' math achievement was significantly related to three factors: (a) the students' report of content coverage, (b) the teacher's level of content knowledge, and (c) prior math ability and classes taken. The researchers also provided two linguistics-related accommodations: dual language tests and linguistically modified items. Unlike the previous studies, this study found no improvement in scores for students who received linguistically modified items.

Datnow, Borman, Stringfield, Overman, and Castellano (2003) studied the effects of comprehensive school reforms on schools with ELL populations. They reported that “the results on implementation and achievement, taken together, suggest that stronger implementations of comprehensive school reform models were in general, associated with better outcomes for LEP (ELL) students” (p. 164). Other research focusing on improving the mathematics achievement of ELL students has focused on achievement grouping (Chang, Singh, & Fuller, 2009), computational skills (Bresser, 2003), computer-based instructional programs (Dixon, 1995), computer usage at home (Kim & Chang, 2010), and project-based learning (Chamot, Dale, O’Malley, & Spanos, 1992).

Summary

As mentioned previously, ELL students perform at or near the bottom of the achievement range nationally (Abedi & Lord, 2001) and in New Mexico (New Mexico Standards Based Assessment, 2010). Brown (2005) reported that assessments that can be considered language-based performance assessments (LBPA) place greater demands on ELL students because the students’ underdeveloped literacy and language skills put them at a distinct disadvantage when attempting to explain answers/solutions in writing. The *New Mexico Standards Based Assessment*, which was used to measure mathematics achievement in the present study, is an LBPA. At the present time, there is no published research on the use of intervention programs or strategies to improve ELL students’ mathematics achievement on language-based performance assessments.

CHAPTER 3

METHODOLOGY

The main purpose of the No Child Left Behind Act of 2001 was to improve the quality of education for all students in the United States. In addition to mandates aimed toward raising student achievement, particularly in reading and math, came mandates requiring greater accountability by states and school districts. These accountability mandates raised the bar for ELL students and held school districts and states accountable for improving the education of ELL students. New Mexico public schools, together with public schools in all states, have struggled to meet this mandate. The purpose of the present study was to investigate the impact of curriculum-based measures on math achievement among all students, with an emphasis on ELL students.

This study addressed the following research questions:

1. What is the impact of curriculum-based measures on the math achievement of ALL students?
2. What is the impact of curriculum-based measures on the math achievement of ELL students?

Design

Collins (1992) was a pioneer in the work on designed experiments in education, where the focus was on investigating how different learning environment designs affect dependent variables in teaching and learning. In discussing methods and designs, Collins, Joseph, and Bielaczyc (2004) argued that designs can be more or less specific, but can never be completely specified and that results can “vary widely depending on things like, participants’ needs, interests, abilities, interpretations, interactions, and goals” (p. 17). They also stated that because educational experiments are carried out in messy situations of actual classrooms, “there are many variables that affect the success of the design, and many of these variables cannot be controlled” (p. 19).

Pituch (2001) proposed a combination large-scale planned variation experimental design and multilevel analysis that would address research questions that went beyond the issue of overall treatment effect. Limitations to treatment effect research include not identifying mechanisms by which the intervention achieves its outcomes (Lipsey, 1988; Pawson & Tilley, 1997; Weiss, 1997), not recognizing that the intervention/treatment works better for some types of individuals than for others (Baron & Kenny, 1986; Petrosino, 2000), and not determining how well the intervention/treatment works across different contexts (Cronbach, 1982; Lipsey & Wilson, 1993; Seltzer, 1994).

The present study employed a quasi-experimental design of experimental and control groups. Analysis of covariance (ANCOVA) tests were applied, one with fourth grade data and one with fifth grade data. The groups consisted of intact classroom groups (more detail below). The research design used in this study was aligned to the planned variation model proposed by Yeh (2000). One strength of the design is that it allows the testing of additional hypotheses along with the main treatment effect. Yeh called this type of study *theory-based evaluation.* This overall design can address whether and how well the intervention worked, who it benefitted, and perhaps the degree to which replication is possible.

One of the potential problems of this type of design is possible confounding treatment effects (Orr, 1999).

Population and Sample

In April of 2008, all third and fourth grade students in eight public elementary schools from one school district in New Mexico were assessed using the *New Mexico Standards Based Assessment*. This is the test used for NCLB reporting, so data were taken from the regular assessment given by the school district (and the state of New Mexico). Data for each of the two grade levels were kept and analyzed separately.

The school district is located in the northwest part of New Mexico and is comprised of city schools, county schools, and schools located on Native lands. Only schools located in the city were used for this investigation. At the time of the study the district had an enrollment of almost 12,000 students, 80 % of whom were Native Americans, 12% were Hispanic, and 7% were Caucasian. Further breakdown of enrollment shows that 75% of the total enrollment was eligible for free/reduced lunch, while 45% were classified as ELL. These numbers were significantly higher than the overall percentages for the state: 50% free/reduced lunch and only 17% ELL (NMPED, 2010). The percentage of both Native American students and students on free/reduced lunch were such that it was not feasible to use these categories, or differentiated categories, as independent variables.

The April 2008 administration of the NMSBA was treated as the pretest in the experimental design in which these third and fourth grade students became fourth and fifth grade students, respectively, the following school year (2008–09). After the administration of the test, four of the eight school principals volunteered their schools for implementation of Curriculum-Based Measures, which were used to supplement the core math curriculum during the 2008–09 school year. These four schools constituted the experimental, or treatment, group and were identified as: Exp 1, Exp 2, Exp 3, and Exp 4 throughout this study. The other four schools used the same core math curriculum, only without the CBMs. These schools constituted the control group and were identified as: Con 1, Con 2, Con 3, and Con 4. The core math curriculum used by all classes and schools in the experimental and control groups had been used the previous year. Designating entire schools as either experimental or control schools was necessary because the designation of treatment groups and control groups within individual schools was considered but rejected by the principals of the schools. Therefore, there was no randomization of subjects or treatment in this study. Neither students nor teachers were informed that the use of CBMs would be part of an educational research investigation. However, the implementation of CBMs in the experimental schools was mandated to the teachers by their respective principals. Scores from the regular administration of the NMSBA in April of 2009 were treated as the posttest scores.

Data from students who had stable enrollment from the time of the pretest through the posttest were analyzed as part of this study. Stable enrollment was defined similarly to the definition required for a student to be a Full Academic Year student under NCLB, namely, the maintenance of continuous enrollment in a given school from day 120 of one academic school year through day 120 of the subsequent school year. Data from students not meeting the above definition of FAY were not included in this study. The assignment of students to specific teachers and classes followed the normal assignment process at each school and was carried out independent of the present study. Information collected for each student included gender, school, grade level, teacher, ELL status, and group (treatment or control), as well as pre- and post-test scores from the NMSBA on reading, writing, science, and math.

The classification of ELL for this study matched the NCLB and NMSBA reporting categories: current, exit, and never. The label *current* ELL, is used for students who have not met English proficiency according to the ACCESS (see chapter 2, section titled "Identification of English Language Learners"). The exit category represents students who have met proficiency on the ACCESS (or NMELPA) within the previous two years. The provisions of NCLB state that a student who has met English proficiency standards still counts for the AYP reporting subgroup of ELL for two years. The ELL category of *never* indicates a student who has never been classified ELL or one who had achieved English proficiency standards more than two years ago.

Tables 3 and 4 compare "pretest" scores for the state of New Mexico, school district, and schools from the district that were included in the investigation. Scores are also categorized to compare the ALL student group and the current ELL group. The scores indicate the percentage of students who are proficient or advanced in reading and math. As can be seen, the district was performing below the levels of the entire state in both reading and math; however, in the fourth grade, one of the control schools (Con 2) and in the fifth grade, one of the experimental schools (Exp 1) did compare favorably to the state levels.

Table 3

Fourth Grade Pretest Proficiency Comparison: State, All District Schools, and Research Schools

	Reading		Math	
Entity	ALL	ELL	ALL	ELL
State	61.0	45.3	54.1	47.0
District	37.2	25.8	31.7	26.1
Con 1	45.6	23.1	33.8	23.1
Con 2	62.8	40.0	51.2	33.3
Con 3	38.1	28.6		
Con 4	31.9	26.7	27.7	33.3
Exp 1	47.7	27.8	34.1	33.3
Exp 2	25.5	9.1	14.9	9.1
Exp 3	29.8	20.0	14.5	10.0
Exp 4	26.1	20.0	17.4	26.7

Note: Numbers indicate percentage of proficient students. Blank spaces indicate samples too small for reporting. Adapted from “NMPED School Fact Sheet,” by New Mexico Public Education Department. Retrieved January 29, 2011, from http://www.ped.state.nm.us/IT/schoolFactSheets.html

Table 4

Fifth Grade Pretest Proficiency Comparison: State, All District Schools, and Research Schools

	Reading		Math	
Entity	ALL	ELL	ALL	ELL
State	51.8	33.3	42.0	27.8
District	31.4	17.1	26.7	20.3
Con 1	46.5	10.0	32.6	20.0
Con 2	34.9	9.1	30.2	18.2
Con 3	44.1	35.3		
Con 4	20.9	18.2	20.9	22.1
Exp 1	50.5	30.8	40.6	23.1
Exp 2	30.2	18.5	11.3	7.4
Exp 3	22.7	20.0	15.2	10.0
Exp 4	28.9	4.8	26.7	9.5

Note: Numbers indicate percentage of proficient students. Blank spaces indicate sample sizes too small for reporting. Adapted from "NMPED School Fact Sheet," by New Mexico Public Education Department. Retrieved January 29, 2011, from http://www.ped.state.nm.us/IT/schoolFactSheets.html

New Mexico Standards Based Assessment (NMSBA)

In compliance with NCLB, each spring the New Mexico Public Education Department (PED) administers the NMSBA in English (in Spanish for students who need it) in the subject areas of reading, writing, math, and science for students in grades 3 through 8 and grade 11. The 11th grade NMSBA also contains a social studies subtest. All of the subtests, with the exception of writing, contain a mixture of multiple choice (MC), short answer (SA), and open-ended (OE) items. All items used on the NMSBA were developed by Pearson with PED overview and aligning to NM standards, and underwent the normal standardized test-item development process. Items were assessed by Pearson test developers and panels of New Mexico educators and then field-tested. The MC items require students to select a correct answer from four alternatives; correct answers to these items are worth 1 point each. The SA items require students to answer questions with either a few words or sentences; possible scores on these items are 0, 1, or 2 points each. The OE items require students to write a paragraph or more and are scored at 0–4 points each. The inclusion of SA and OE items makes the NMSBA a language-based performance assessment.

The fourth grade NMSBA math subtest consisted of 57 total items (40 MC, 15 SA, and 2 OE items) and 78 possible points (MC = 1 @ 40 = 40 points; SA = 2 @ 15 = 30 points; OE = 4 @ 2 = 8 points). The fifth grade NMSBA math subtest consisted of 62 total items (43 MC, 16 SA, and 3 OE items) and 87 possible points (MC = 1 @ 43 = 43 points; SA = 2 @ 16 = 32 points; OE = 4 @ 2 = 8 points). The technical report states that all NMSBA student scores are "scaled using item response theory and scores are reported as scaled scores, which is the predominant reporting method used on most large-scale educational assessments in the U.S." (*New Mexico Standards-Based Assessment Technical Manual,* 2010, p. 43). May, Perez-Johnson, Haimson, Sattar, and Gleason (2009) stated that because scaled scores are measured on a continuous scale, they provide greater precision than proficiency level scores, which are measured on an ordinal scale.

May et al. (2009) argued that a researcher must be able to show that a test is both reliable and valid to justify its use as a dependent measure in educational experiments designed to ascertain treatment effects. They defined test reliability "as the degree to which the state assessment provides scores that are sufficiently free of random measurement that they can be used to detect program effects" (p. 5). A more general description of test reliability refers to the degree of consistency obtained between different administrations of a particular test. The NMSBA technical manual reports that the coefficient alpha (∞) is used to report the reliability of the NMSBA. The manual defines the coefficient alpha as the "proportion of variance attributable to examinees' true abilities divided by the observed variance in the test scores" (*New Mexico Standards-Based Assessment Technical Manual*, 2010, p. 51). Reliability scores computed with the coefficient alpha can range from ∞ = 0.00 to 1.00, where a score of ∞ = 0.00 means there is no reliability and ∞ = 1.00 means perfect reliability. The reliability score converted to a percent indicates the percentage of the score that is accurate while the remaining percentage indicates the amount of standard error. The calculated reliability coefficients as reported for demographic subgroups range from ∞ = 0.87 to 0.93. Because these reported figures are greater than ∞ = 0.85, Pearson considers the NMSBA to be reliable for this type of test (*New Mexico Standards-Based Assessment Technical Manual*, 2010, p. 53).

Because the NMSBA is also used to inform parents, teachers, and the general public about student achievement, scaled scores need to be associated with the performance standards in a comprehensible way. To accomplish this goal, Pearson and PED created a four-level classification or performance system. The four performance levels are Beginning Step, Nearing Proficiency, Proficient, and Advanced. Reliability must be established for these performance levels as well; it is measured by the accuracy and consistency of classifying students in these performance levels. Again, the technical manual (*New Mexico Standards-Based Assessment Technical Manual*, 2010) reports that the coefficient alpha (∞) is used for this purpose. Calculated accuracy and consistency scores range from ∞ = 0.88 (5th grade Nearing Proficient/Proficient) to ∞ = 0.97 (4th grade Proficient/Advanced). Thus, both the fourth and fifth grade NMSBA math tests have accurate and consistent reliability on performance levels.

May et al. (2009) defined test validity as "the degree to which the state assessment adequately measures the outcomes targeted by the intervention" (p. 5). A more general definition of test validity answers the question, "Does the test measure what it is intended to measure?" The technical manual addresses three types of validity for the NMSBA: test content (content validity), internal structure (construct validity), and relationship to other variables (concurrent validity). The technical manual reports that by following the accepted protocol of item development and by ensuring that the items are aligned to the NM standards, the NMSBA is a valid assessment instrument. Internal structure refers to the fact that a collection of items should be used to measure the same construct. Correlation coefficients were calculated for all items intended to measure each construct. These scores ranged from ∞ = 0.69 to 1.00 (on a scale of 0–1.0), indicating that the items have acceptable levels of correlation, and that each set of items is measuring a given discrete construct.

In discussing validity, in terms of relationships to other variables (concurrent validity), the technical manual states that an inability to read will affect performance on other subtests. Correlations were calculated among NMSBA subject tests to determine concurrent validity. The conclusion in the technical manual is that "on average the tests share roughly 40% of the variance across subjects. . . . This suggests that while a common underlying trait is likely shared, the NMSBA tests are uniquely measuring subject-specific constructs as intended" (*New Mexico Standards-Based Assessment Technical Manual*, 2010, pp. 65-66).

Data Analysis

This study compared students' math achievement scores on the NMSBA to determine whether there were significant differences between achievement scores between students who used weekly curriculum-based measures (experimental group) and those who did not (control group). The post-treatment scores for each group were compared to determine whether significant differences existed as a result of the treatment. Data gathered from this research process were collected and entered into a statistics software package—Statistical Package for the Social Sciences (SPSS). SPSS was used for all statistical analyses and the significance was set at the .05 level for all inferential tests.

Pretest scores (as defined earlier) were used as the covariate in each ANCOVA analysis, one for the fourth grade and one for the fifth grade. Use of baseline measures has been shown to increase statistical power when they are used as covariates in impact analyses (Bloom, Richburg-Hayes, & Black, 2007; Shandish, Cook, & Campbell, 2002).

New Mexico Math Curriculum

The New Mexico Public Education Department developed the New Mexico math content standards, benchmarks, and performance standards to "establish an articulated, coordinated, and comprehensive description of the content and skills students should learn at specific grade levels in the study of mathematics" (*Complete K-12 Math Standards Document*, 2008, p. 2). The curriculum identifies what students should know and be able to do across grade levels. The overall math curriculum for both the fourth and fifth grades is established on five strands: number and operations (N), algebra (A), geometry (G), measurement (M), and data and probability (D). Each strand has associated benchmarks and performance standards, which state in greater detail what students should know and be able to do. An example of a fourth grade performance standard is "4.N.1.4 Recognizes classes of numbers (e.g., odd, even, factors, multiples, square numbers) and apply these concepts in problem solving situations" (p. 12).

In addition to the math content standards, NMPED also provided a list of approved publishers from which schools and/or districts could purchase core math materials (*Core/Basal Instructional Material Adoption Approvals, 2006*). All of the schools in this research study used Harcourt School Publishers as their core math curriculum.

Curriculum-Based Measures (CBMs)

In addition to all of the resources that accompanied the Harcourt math program, the experimental schools were provided with an independently created weekly assessment called a *curriculum-based measure.* For the sake of elementary students, they were given the name "Math Monsters." Individual questions for the Math Monsters were developed and identified by New Mexico math performance standards. Each Math Monster had eight questions and a total of 10 possible points. Six of the questions were multiple choice and worth 1 point each, and two questions were constructed response-type questions and worth 2 points each (6 multiple choice * 1 point and 2 constructed response @ 2 points = 10 total points). Each Math Monster had at least one question from each of the performance strands [number and operations (1), algebra (2), geometry (3), measurement (4), and data and probability (5)]. Sample Math Monsters for both fourth and fifth grades are located in Appendices A and B, respectively.

Summary

A quantitative methodology, following that proposed by Yeh (2000), was used to gather demographic and achievement data on fourth and fifth grade public school students in one New Mexico school district. The primary purpose of this research study was to assess the impact of CBMs on math achievement. Using students' math scores from the 2009 NMSBA administration as baseline pretest scores, the study then used the 2010 NMSBA math scores as the experimental measure to determine whether use of weekly curriculum-based measures produced significant results. Other factors were also assessed to determine other possible effects and interactions. The analysis allowed for the development of conclusions and generalizations that are reported in the next chapter.

CHAPTER 4

RESULTS

The purpose of this research study was to determine the impact of weekly curriculum-based measures on elementary students' math achievement on a language-based performance assessment. The impact of the measures on ELL students' achievement was of special interest. At the current time, there is no published research on the use of intervention programs or strategies to improve ELL students' mathematics achievement on LBPAs.

Fourth Grade Results

A Pearson Product-Moment Correlation test revealed a significant positive correlation between the NMSBA math pre- and post-test scores ($r < .779$, $p < .000$). This coefficient is within the range of expectations for pretests and posttests.

Table 5

Fourth Grade Mean Pretest Scores: Control and Treatment Groups

Treatment	Mean	*SD*	*N*
Control	602.43	26.381	139
Experimental	593.89	23.305	144

There was a significant difference between the treatment groups on the pretest in favor of the control group [ANOVA $F(1,281) = 8.350$, $p < .004$ partial

$\eta^2 = .029$] (see Table 5). A test of equality of error variances showed that the two groups were homogeneous [Levene's $F(1, 281) = 3.692$, $p < .056$].

Table 6

Fourth Grade Posttest Scores: Treatment, Gender, and ELL

Treatment	Gender	ELL Status	Mean	*SD*	*N*
Control	F	Current	627.33	31.342	3
		Exit	631.20	32.836	10
		Never	630.23	33.890	53
	M	Current	617.05	30.462	20
		Exit	629.14	31.567	7
		Never	628.30	31.932	46
Experimental	F	Current	621.75	21.225	8
		Exit	635.56	30.924	9
		Never	623.03	31.811	39
	M	Current	599.37	27.840	32
		Exit	638.20	36.657	5
		Never	623.51	32.424	51

Table 7

Fourth Grade Analysis of Covariance: Treatment, Gender, and ELL

Source	*SS*	*df*	*MS*	*F*	*p*	Partial η^2
Covariate (Pretest)	162859.787	1	162859.787	402.480	.000	.599
Treatment (T)	106.511	1	106.511	.263	.608	.001
Gender (G)	204.177	1	204.177	.505	.478	.002
ELL (E)	780.197	2	390.099	.964	.383	.007
T x G	469.949	1	469.949	1.161	.282	.004
T x E	1075.144	2	537.572	1.329	.267	.010
G x E	973.701	2	486.850	1.203	.302	.009
T x G x E	1023.498	2	511.749	1.265	.284	.009
Error	109253.051	270	404.641			

To measure the effects of the treatment, a series of analysis of covariance tests was applied to the data. The pretest scores served as the covariate and the posttest scores served as the dependent variable in each case. In the first ANCOVA analysis on fourth grade data, there were three independent variables: treatment (experimental, control), gender, and ELL (current, exit, never). Means, standard deviations, and group and cell sizes are reported in Table 6. The results of the ANCOVA analysis, as displayed in Table 7, show that there were no significant main effects for treatment, gender, or ELL ($p < .05$). Further, there were no significant effects for any of the four interactions in this model ($p < .05$). The effect sizes for all main and interaction effects were small (partial $\eta^2 < .011$). Of the 12 cells in this model, one contained only three cases (see Table 6), which is less than the recommended number for interaction analysis ($n > 4$). The results should be interpreted with a degree of caution due to this anomaly, although in this case, the absence of any significant interactions renders it of little concern.

To further examine the data, I employed a paired *t*-test on pretest and posttest scores to check for differences. First, the results showed a significant increase between the pretest (M = 598.08, SD = 25.185) and posttest (M = 623.34, SD = 32.524), for a mean difference of 25.25 (t = 20.744, df = 282, $p < .000$). A breakdown between groups shows a significant increase for the control group between the pretest (M = 603.43, SD = 26.381) and posttest (M = 627.65, SD = 32.263), for a mean difference of 25.22 (t = 15.633, df = 138, $p < .000$). The pretest-posttest difference for the experimental group was nearly identical: pretest (M = 593.89, SD = 23.305) and posttest (M = 619.18, SD = 32.343), for a mean difference of 25.29 (t = 13.879, df = 143, $p < .000$). As a further check, I compared the pretest-posttest gains scores between the control and experimental groups using a one-way factorial ANOVA. Because the gain scores were almost identical, there was no significant difference between groups, and no measurable effect size [$F(1, 281) = .001$, $p < .975$, partial $\eta^2 = .000$].

Table 8

Fourth Grade Pre- and Posttest Scores: Schools and Gender

		Pretest			Posttest		
School	Gender	Mean	*SD*	*N*	Mean	*SD*	*N*
Control 1	F	603.45	24.968	22	627.41	28.872	22
	M	595.50	24.958	26	623.08	34.080	26
	Total	599.15	25.018	48	625.06	31.544	48
Control 2	F	624.00	27.126	14	647.57	39.420	14
	M	599.94	37.763	17	628.41	36.941	17
	Total	610.81	35.037	31	637.06	38.667	31
Control 3	F	595.15	22.345	13	615.92	30.623	13
	M	599.29	19.839	14	620.14	24.516	14
	Total	597.30	20.777	27	618.11	27.165	27
Control 4	F	601.59	21.734	17	630.59	31.042	17
	M	605.63	22.262	16	630.13	27.873	16
	Total	603.55	21.742	33	630.36	29.087	33
Control Total	F	605.70	25.642	66	630.24	33.149	66
	M	599.48	26.864	73	625.30	31.485	73
	Total	602.43	26.381	139	627.65	32.263	139
Exp. 1	F	599.92	17.307	12	643.83	22.225	12
	M	601.18	21.119	17	641.18	24.123	17
	Total	600.66	19.312	29	642.28	22.986	29
Exp. 2	F	595.58	18.143	12	618.67	18.465	12
	M	590.15	25.819	26	601.23	36.377	26
	Total	591.87	23.555	38	606.74	32.603	38
Exp. 3	F	598.94	25.638	17	628.88	41.694	17
	M	590.28	25.672	29	620.41	31.795	29
	Total	593.48	25.723	46	623.54	35.555	46
Exp. 4	F	596.00	23.047	15	610.07	19.196	15
	M	585.63	21.341	16	602.88	19.473	16
	Total	590.65	22.436	31	606.35	19.362	31

Table 8 (continued)

School	Gender	Pretest Mean	Pretest *SD*	Pretest *N*	Posttest Mean	Posttest *SD*	Posttest *N*
Exp. Total	F	597.64	21.342	56	624.86	30.300	56
	M	591.50	24.288	88	615.57	33.242	88
	Total	593.89	23.305	144	619.18	32.343	144
Grand Total	F	602.00	24.010	122	627.77	31.857	122
	M	595.12	25.718	161	619.98	32.720	161
	Total	598.08	25.185	283	623.34	32.524	283

Table 9

Fourth Grade Analysis of Covariance: Schools and Gender

Source	*SS*	*df*	*MS*	*F*	*p*	Partial η^2
Covariate (Pretest)	153368.276	1	153368.276	408.410	.000	.606
School (S)	14578.864	7	2082.695	5.546	.000	.127
Gender (G)	93.211	1	93.211	.248	.619	.001
S x G	1750.849	7	250.121	.666	.701	.017
Error	99889.841	266	375.526			

In the second ANCOVA analysis for the fourth grade, there were two independent variables: school and gender. The original plan called for three independent variables: school, gender, and ELL, but too many cells had a smaller number of cases than required for the interaction analysis ($n < 4$). Pretest and posttest means, standard deviations, and group and cell sizes are reported in Table

8. The results of the ANCOVA analysis, as displayed in Table 9, show a significant main effect for school, with a relatively large effect size [$F(1,7) = 5.546$, $p = .000$, partial $\eta^2 = .127$]. Both the main effect for gender and for the interaction were not significant ($p < .05$).

To further examine the results by school, I employed a one-way factorial ANOVA and Tukey's post hoc analysis on pretest scores. There was a significant difference between individual schools on the pretest [ANOVA $F(1,7) = 2.431$, $p < .020$, partial $\eta^2 = .058$]. A test of equality of error variances showed that the two groups were homogeneous [Levene's $F(7, 275) = 1.793$, $p < .089$]. Tukey's HSD post hoc analysis revealed that control school 2 had pretest scores significantly higher than experimental school 2 (*MD* = 18.94, $p < .037$) and experimental school 4 (*MD* = 20.16, $p < .032$).

I also employed a one-way factorial ANOVA and Tukey's post hoc analysis on math gain scores (posttest minus pretest). There was a significant difference between the schools on gain scores [ANOVA $F(1,7) = 6.333$, $p < .000$, partial $\eta^2 = .139$]. A test of equality of error variances showed that the two groups were homogeneous [Levene's $F(7, 275) = 1.536$, $p < .155$].

On math gain scores by school, Tukey's HSD post hoc analysis revealed some significant differences. Experimental school 1 had significantly higher gain scores than five other schools: control school 1 (*MD* = 15.70, $p < .014$), control school 2 (*MD* = 15.36, $p < .048$), control school 3 (*MD* = 20.81, $p < .002$), experimental school 2 (*MD* = 26.75, $p < .000$), and experimental school 4 (*MD* = 25.91, $p < .000$). Experimental school 3 had significantly higher gain scores than two other schools: experimental school 2 (*MD* = 15.20, $p < .000$), and experimental school 4 (*MD* = 14.36, $p < .032$).

Table 10

Fourth Grade Pre- and Posttest Scores: Teachers and Gender

School/ Teacher/	Gender	Pretest Mean	Pretest *SD*	Pretest *N*	Posttest Mean	Posttest *SD*	Posttest *N*
C1 /T1	F	600.33	21.963	15	623.07	24.976	15
	M	592.61	24.888	18	622.72	32.947	18
	Total	596.12	23.566	33	622.88	29.148	33
C1/T2	F	610.14	31.302	7	636.71	36.248	7
	M	602.00	25.506	8	623.88	38.868	8
	Total	605.80	27.620	15	629.87	36.911	15
C2/T1	F	617.00	29.282	8	636.00	34.372	8
	M	613.57	21.314	7	644.14	16.088	7
	Total	615.40	25.031	15	639.80	26.820	15
C2/T2	F	633.33	23.019	6	663.00	43.433	6
	M	590.40	44.573	10	617.40	43.889	10
	Total	606.50	42.772	16	634.50	48.004	16
C3/T1	F	585.25	18.328	8	607.75	29.154	8
	M	612.67	10.367	6	638.00	18.804	6
	Total	597.00	20.505	14	620.71	28.896	14
C3/T2	F	611.00	20.00	5	629.00	31.281	5
	M	589.25	19.638	8	606.75	19.631	8
	Total	597.62	21.900	13	615.31	26.036	13
C4/T1	F	592.14	22.858	7	630.14	30.716	7
	M	600.57	20.935	7	627.00	33.232	7
	Total	596.36	21.507	14	628.57	30.786	14

Table 10 (continued)

School/ Teacher	Gender	Pretest Mean	*SD*	*N*	Posttest Mean	*SD*	*N*
C4/T2	F	608.20	19.326	10	630.90	32.922	10
	M	609.56	23.681	9	632.56	24.764	9
	Total	608.84	20.892	19	631.68	28.552	19
Con Total	F	605.70	25.642	66	630.24	33.149	66
	M	599.48	26.864	73	625.30	31.485	73
	Total	602.43	26.381	139	627.65	32.263	139
E1 /T1	F	599.33	15.403	9	643.56	24.976	9
	M	602.50	12.276	6	638.67	23.364	6
	Total	600.60	13.855	15	641.60	23.612	15
E1/T2	F	601.67	26.312	3	644.67	14.844	3
	M	600.45	25.232	11	642.55	25.540	11
	Total	600.71	24.424	14	643.00	23.163	14
E2/T1	F	602.00	13.914	6	614.00	7.294	6
	M	592.92	26.983	12	609.50	36.995	12
	Total	595.94	23.398	18	611.00	30.100	18
E2/T2	F	589.17	20.779	6	623.33	25.390	6
	M	587.79	25.550	14	594.14	35.630	14
	Total	588.20	23.679	20	602.90	35.023	20
E3/T1	F	594.50	13.867	8	617.63	24.295	8
	M	588.89	24.374	9	616.33	28.644	9
	Total	591.53	19.736	17	616.94	25.864	17
E3/T2	F	591.33	10.066	3	657.67	4.163	3
	M	588.00	24.052	9	626.89	36.628	9
	Total	588.83	21.010	12	634.58	34.244	12
E3/T3	F	608.67	40.173	6	629.50	63.475	6
	M	593.27	29.850	11	618.45	32.309	11
	Total	598.71	33.447	17	622.35	44.058	17

Table 10 (continued)

School/ Teacher	Gender	Pretest Mean	Pretest SD	Pretest N	Posttest Mean	Posttest SD	Posttest N
E4/T1	F	599.67	29.924	6	606.83	15.368	6
	M	586.00	22.192	9	600.78	22.348	9
	Total	591.47	25.481	15	603.20	19.472	15
E4/T2	F	593.56	18.789	9	612.22	22.004	9
	M	585.14	21.943	7	605.57	16.339	7
	Total	589.87	19.986	16	609.31	19.407	16
Exp. Total	F	597.64	21.342	56	624.86	30.300	56
	M	591.50	24.288	88	615.57	33.242	88
	Total	593.89	23.305	144	619.18	32.343	144
Grand Total	F	602.00	24.010	122	627.77	31.857	122
	M	595.12	25.718	161	619.98	32.720	161
	Total	598.08	25.185	283	623.34	32.524	283

Note: School/Teacher identifies teachers with their respective schools (C1 = Control school 1).

Table 11

Fourth Grade Analysis of Covariance: Teachers and Gender

Source	*SS*	*df*	*MS*	*F*	*p*	Partial η^2
Covariate (Pretest)	146902.941	1	146902.941	409.162	.000	.623
Teacher (T)	20358.501	16	1272.406	3.544	.000	.186
Gender (G)	334.756	1	334.756	.932	.335	.004
T x G	7067.178	16	441.699	1.230	.245	.074
Error	99889.841	266	375.526			

In the third and final ANCOVA analysis on fourth grade data, there were two independent variables: teacher and gender. The original plan called for three independent variables: teacher, gender, and ELL status; however, too many of the ELL cells had a small number of cases ($n < 4$), so ELL status was not included as a main effect (independent variable) in this analysis. Within this analysis, 2 of the 34 cells contained a small number of cases (see Table 10) and results should be taken with some degree of caution. Pretest and posttest means, standard deviations, and group and cell sizes are reported in Table 10. The results of the ANCOVA analysis, as displayed in Table 11, show a significant main effect for teacher, with a substantial effect size [$F(1, 16) = 3.544$, $p = .000$, partial $\eta^2 = .186$]. There was no significant main effect for gender or interaction ($p < .05$).

Examining teacher results further, I employed a one-way factorial ANOVA and Tukey's post hoc analysis on pretest scores. There were no significant differences between teachers on the pretest [ANOVA $F(1,16) = 1.569$, $p < .078$, partial $\eta^2 = .092$]. A test of equality of error variances showed that the two groups were homogeneous [Levene's $F(33, 249) = 1.151$, $p < .270$].

I also employed a one-way factorial ANOVA and Tukey's post hoc analysis on math gain scores (posttest minus pretest) by teacher. There was a significant difference among teachers on gain scores [ANOVA $F(1,16) = 3.807$, $p < .000$, partial $\eta^2 = .186$]. A test of equality of error variances showed that the two groups were homogeneous [Levene's $F(16, 266) = 1.321$, $p < .184$].

On math gain scores (posttest minus pretest) by teacher, Tukey's HSD post hoc analysis revealed some significant differences. There was only one significant difference in gain scores across treatment: teacher E3T2 had significantly higher gain scores than teacher C4T2 (*MD* = 28.06, $p < .027$). All other significant differences occurred between teachers at experimental schools. There were no significant differences in gain scores between teachers at the same school. In addition to having significantly higher gain scores than a control school teacher, teacher E3T2 also had significantly higher gain scores than four experimental teachers: E2T1 (*MD* = 30.69, $p < .002$), E2T2 (*MD* = 31.05, $p < .001$), E4T1 (*MD* = 34.02, $p < .001$), and E4T2 (*MD* = 26.31, $p < .033$).

Both teachers at experimental school 1 (E1T1 and E1T2) had significantly higher gain scores than the same three teachers: E2T1, E2T2, and E4T1. Results for teacher E1T1 are: E2T1 (*MD* = 26.30, $p < .007$), E2T2 (*MD* = 25.94, $p < .013$), and E4T1 (*MD* = 29.27, $p < .004$). Results for teacher E1T2 are: E2T1 (*MD* = 27.23, $p < .008$), E2T2 (*MD* = 27.59, $p < .005$), and E4T1 (*MD* = 30.55, $p < .003$).

Fifth Grade Results

A Pearson Product-Moment Correlation test revealed a significant positive correlation between the NMSBA math pretest and posttest scores ($r < .823$, $p < .000$). This coefficient is within the range of expectations for pretests and posttests.

Table 12

Fifth Grade Mean Pretest Scores: Control and Treatment Groups

Treatment	Mean	SD	*N*
Control	621.86	24.714	139
Experimental	613.38	26.398	144

There was a significant difference between the treatment groups on the pretest in favor of the control group (see Table 12) [ANOVA $F(1,292) = 8.023$, $p < .005$, partial $\eta^2 = .027$]. A test of equality of error variances showed that the two groups were homogeneous: [Levene's $F(1, 292) = 0.772$, $p < .380$].

Table 13

Fifth Grade Pre- and Posttest Scores: Treatment, Gender, and ELL

			Pretest			Posttest		
Treat	G	ELL	Mean	*SD*	*N*	Mean	*SD*	*N*
Con	F	Current	617.47	18.738	15	630.07	25.181	15
		Exit	623.40	14.082	5	659.00	6.205	5
		Never	623.33	24.156	55	661.58	29.867	55
		Total	622.16	22.551	75	655.11	30.533	75
	M	Current	612.67	23.003	21	633.43	32.400	21
		Exit	608.25	26.424	4	623.50	38.854	4
		Never	627.62	28.173	39	654.31	35.763	39
		Total	621.50	27.208	64	645.53	36.110	64

Table 13 (continued)

Treat	G	ELL	Pretest Mean	Pretest *SD*	Pretest *N*	Posttest Mean	Posttest *SD*	Posttest *N*
Con	Total	Current	614.67	21.180	36	632.03	29.263	36
		Exit	616.67	20.609	9	643.22	30.585	9
		Never	625.11	25.839	94	658.56	32.459	94
		Total	621.86	24.714	139	650.70	33.439	139
Exp	F	Current	602.23	17.096	22	628.73	19.533	22
		Exit	643.87	29.054	8	667.75	34.964	8
		Never	615.38	26.168	37	645.54	26.103	37
		Total	614.46	26.669	67	642.67	27.720	67
	M	Current	608.00	24.346	35	630.57	27.192	35
		Exit	653.00	23.338	4	662.50	34.337	4
		Never	612.51	25.487	49	639.45	30.196	49
		Total	612.56	26.313	88	636.97	29.708	88
Exp	Total	Current	605.77	21.852	57	629.86	24.348	57
		Exit	646.92	26.569	12	666.00	33.259	12
		Never	613.74	25.669	86	642.07	28.507	86
		Total	613.38	26.398	155	639.43	28.913	155
Total	F	Current	608.41	19.094	37	629.27	21.670	37
		Exit	636.00	25.807	13	664.38	27.305	13
		Never	620.13	25.150	92	655.13	29.351	92
		Total	618.53	24.793	142	649.24	29.795	142
	M	Current	609.75	23.749	56	631.64	28.996	56
		Exit	630.63	33.239	8	643.00	39.835	8
		Never	619.20	27.605	88	646.03	33.419	88
		Total	616.32	26.970	152	640.57	32.719	152
Grand Total		Current	609.22	21.915	93	630.70	26.225	93
		Exit	633.95	28.168	21	656.24	33.406	21
		Never	619.68	26.309	180	650.68	31.647	180
		Total	617.39	25.921	294	644.76	31.587	294

Note: *Treat.* refers to treatment group (Control or Experimental.). *G* stands for gender.

Table 14

Fifth Grade Analysis of Covariance: Treatment, Gender, and ELL

Source	*SS*	*df*	*MS*	*F*	*p*	Partial η²
Covariate (Pretest)	146711.338	1	146711.338	419.986	.000	.599
Treatment (T)	58.472	1	58.472	.167	.683	.001
Gender (G)	1876.804	1	1876.804	5.373	.021	.019
ELL (E)	6773.232	2	3386.616	9.695	.000	.065
T x G	15.670	1	15.670	.045	.832	.000
T x E	2159.445	2	1079.722	3.091	.047	.022
G x E	2091.172	2	1045.596	2.993	.052	.021
T x G x E	1352.130	2	676.065	1.935	.146	.014
Error	98160.174	281	349.324			

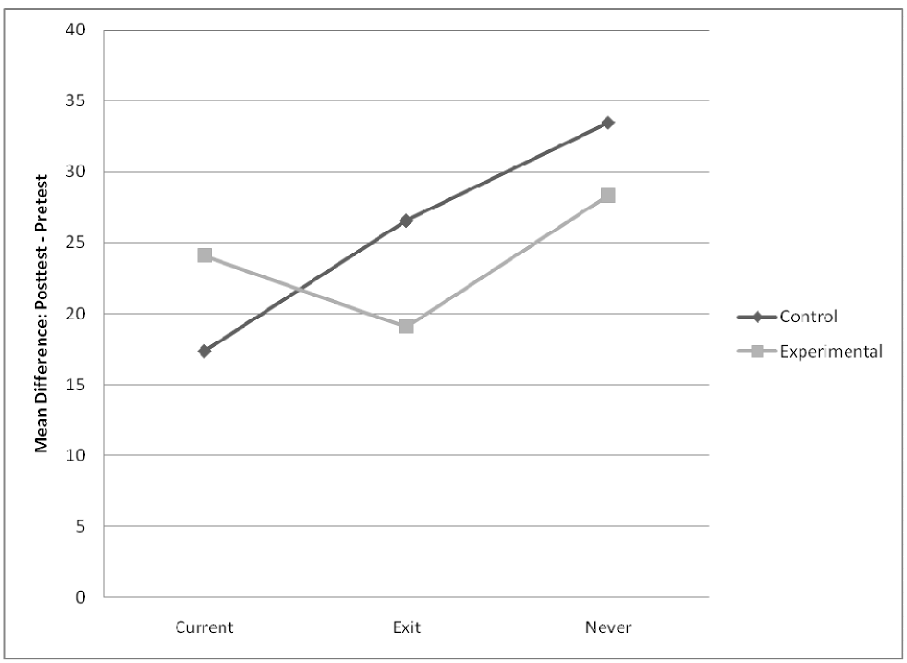


Figure 1. Pre-posttest mean difference as a function of treatment (control, experimental) and ELL (current, exit, never)

To measure the effects of the treatment on fifth grade subjects, a series of analysis of covariance tests (ANCOVA) were applied to the data. The pretest scores served as the covariate and posttest scores served as the dependent variable in each case. In the first ANCOVA analysis, there were three independent variables: treatment (experimental, control), gender, and ELL (current, exit, never). Pretest and posttest means, standard deviations, and group and cell sizes are reported in Table 13.

The results of the ANCOVA analysis, as displayed in Table 14, show that there was no main effect for treatment; however, significant main effects were found for gender [$F(1,281) = 5.373$, $p < .021$, partial $\eta^2 = .019$], and for ELL [$F(2,281) = 9.695$, $p < .000$, partial $\eta^2 = .065$]. The results also revealed a significant interaction between treatment and ELL (T x E), $F(2, 281) = 3.091$, $p < .047$, partial $\eta^2 = .022$] (see Figure 1). Current ELL students in the experimental group had higher scores than current ELL students in the control group. However, the control exit and control never groups had higher scores than their experimental counterparts. There were no other significant interaction effects between treatment and gender (T x G), gender and ELL (G x E), or the 3-way interaction among treatment, gender, and ELL (T x G x E). The effect sizes for the main effect of gender and the significant interaction were small, but the effect size for the main effect of ELL was moderate (partial $\eta^2 = .065$).

To further analyze the data, I employed a one-way analysis of variance (ANOVA) on the pretest data and found a significant difference between the ELL groups [$F(2,291) = 10.218$, $p = .000$, partial $\eta^2 = .066$]. Tukey's HSD post hoc analysis revealed significant differences between all three pairs of ELL groups: current-exit (*MD* = 24.74, p =.000), current-never (*MD* = 10.46, $p < .004$), and exit-never (*MD* = 14.27, $p < .038$). However, an ANOVA revealed no significant difference for gender on the pretest [$F(1,291) = .531$, $p < .467$, partial $\eta^2 = .002$].

Table 15

Fifth Grade Pre- and Posttest Scores: Schools and Gender

		Pretest			Posttest		
School	G	Mean	*SD*	*N*	Mean	*SD*	*N*
Con 1	F	621.24	24.289	25	661.64	31.079	25
	M	624.27	27.091	15	651.73	31.687	15
	Total	622.37	25.074	40	657.93	31.280	40
Con 2	F	624.80	25.904	20	659.60	24.676	20
	M	620.17	32.806	18	649.39	39.530	18
	Total	622.61	29.061	38	654.76	32.516	38
Con 3	F	624.85	22.880	13	661.15	36.104	13
	M	629.47	25.318	17	653.94	36.986	17
	Total	627.47	23.994	30	657.07	36.157	30
Con 4	F	618.35	15.696	17	635.59	25.217	17
	M	610.57	19.437	14	623.71	28.995	14
	Total	614.84	17.624	31	630.23	27.194	31
Con Total	F	622.16	22.551	75	655.11	30.533	75
	M	621.50	27.208	64	645.53	36.110	64
	Total	621.86	24.714	139	650.70	33.439	139
Exp. 1	F	634.40	39.371	10	657.00	41.425	10
	M	631.67	30.266	12	652.75	24.757	12
	Total	632.91	33.854	22	654.68	32.576	22

Table 15 (continued)

School	G	Pretest			Posttest		
		Mean	*SD*	*N*	Mean	*SD*	*N*
Exp. 2	F	597.88	18.624	17	639.35	26.232	17
	M	604.46	29.149	28	637.82	35.203	28
	Total	601.98	25.650	45	638.40	31.800	45
Exp. 3	F	605.79	16.744	24	633.75	22.097	24
	M	606.90	18.792	30	629.93	29.294	30
	Total	606.41	17.754	54	631.63	26.175	54
Exp. 4	F	632.63	18.733	16	650.63	22.899	16
	M	621.83	22.033	18	636.83	20.583	18
	Total	626.91	20.964	34	643.32	22.482	34
Exp Total	F	614.46	26.669	67	642.67	27.720	67
	M	612.56	26.313	88	636.97	29.708	88
	Total	613.38	26.398	155	639.43	28.913	155
Grand Total	F	618.53	24.793	142	649.24	29.795	142
	M	616.32	26.970	152	640.57	32.719	152
	Total	617.39	25.921	294	644.76	31.587	294

Table 16

Fifth Grade Analysis of Covariance: Schools and Gender

Source	*SS*	*df*	*MS*	*F*	*p*	Partial η^2
Covariate (Pretest)	159713.783	1	159713.783	480.196	.000	.634
School (S)	15981.478	7	2283.068	6.864	.000	.148
Gender (G)	2860.206	1	2860.206	8.600	.004	.030
S x G	917.623	7	131.089	.394	.905	.010
Error	92130.453	277	332.601			

In the second ANCOVA analysis on fifth grade data, there were two independent variables: school and gender. The original plan was to test three independent variables, including ELL; however, too many of the cells had a smaller number of cases than required ($n < 4$), so ELL was omitted from this particular analysis. Pretest and posttest means, standard deviations, and group and cell sizes are reported in Table 15. The results of the ANCOVA analysis, as displayed in Table 16, show significant main effects for school [$F(1,7) = 6.864$, $p < .000$, partial $\eta^2 = .148$], with sizable effect size, and gender [$F(1,1) = 8.600$, $p < .004$, partial $\eta^2 = .030$], but no significant interaction between the two variables.

To examine the fifth grade school results further, I employed a one-way factorial ANOVA. Results showed a significant difference among schools on the pretest [ANOVA $F(1,7) = 7.620$, $p < .000$, partial $\eta^2 = .161$]. The significant differences and associated effect size should be interpreted cautiously because a test of equality of variances showed that the subjects grouped by school were not homogeneous [Levene's $F(15, 278) = 2.538$, $p < .001$]. Tukey's HSD post hoc analysis revealed a pattern of significant differences in pretest scores: experimental school 2 was significantly lower than control school 1 ($MD = 20.40$, $p < .003$), control school 2 ($MD = 20.63$, $p < .003$), control school 3 ($MD = 25.49$, $p < .000$), experimental school 1 ($MD = 30.93$, $p < .000$), and experimental school 4 ($MD = 24.93$, $p < .000$). Similarly, experimental school 3 scored significantly lower than control school 1 ($MD = 15.97$, $p < .037$), control school 2 ($MD =$

16.20, $p < .037$), control school 3 (*MD* = 21.06, $p < .004$), experimental school 1 (*MD* = 26.50, $p < .001$), and experimental school 4 (*MD* = 20.50, $p < .003$).

A one-way ANOVA revealed no significant difference between boys and girls on the pretest: [ANOVA $F(1, 292) = .531$, $p < .467$]. A test of equality of variances showed that the subjects grouped by gender were homogeneous [Levene's $F(1, 292) = .787$, $p < .376$].

I also employed a one-way factorial ANOVA and Tukey's post hoc analysis on math gain scores (posttest minus pretest). There was a significant difference between the schools on gain scores [ANOVA $F(1,7) = 7.153$, $p < .000$, partial $\eta^2 = .149$]. A test of equality of error variances showed that the two groups were homogeneous [Levene's $F(7, 286) = 1.260$, $p < .270$].

On math gain scores (posttest minus pretest) by school, Tukey's HSD post hoc analysis revealed some significant differences. Control school 4 scored significantly lower than two other control schools, control school 1 (*MD* = 20.16, $p < .000$) and control school 2 (*MD* = 16.77, $p < .005$). Control school 4 also scored significantly lower than experimental school 2 (*MD* = 21.04, $p < .000$). Experimental school 4 also scored lower than control school 1 (*MD* = 19.14, $p < .000$), control school 2 (*MD* = 15.75, $p < .008$), and experimental school 2 (*MD* = 20.01, $p < .000$). Finally, experimental school 2 scored significantly higher than experimental school 1 (*MD* = 14.65, $p < .048$).

Table 17

Fifth Grade Pre- and Posttest Scores: Teachers and Gender

School/ Teacher	Gender	Pretest			Posttest		
		Mean	*SD*	*N*	Mean	*SD*	*N*
C1 /T1	F	619.38	23.340	13	659.54	35.496	13
	M	618.57	29.342	7	644.00	30.332	7
	Total	619.10	24.821	20	654.10	33.824	20
C1/T2	F	623.25	26.161	12	663.92	26.872	12
	M	629.25	25.866	8	658.50	33.278	8
	Total	625.65	25.531	20	661.75	28.870	20
C2/T1	F	624.70	32.301	10	667.00	26.449	10
	M	627.80	34.257	10	667.00	36.869	10
	Total	626.25	32.444	20	667.00	31.229	20
C2/T2	F	624.90	19.319	10	652.20	21.545	10
	M	610.63	30.294	8	627.38	32.408	8
	Total	618.56	25.075	18	641.17	28.971	18
C3/T1	F	622.33	23.022	9	651.89	30.702	9
	M	636.40	18.119	5	658.60	33.598	5
	Total	627.36	21.820	14	654.29	30.635	14
C3/T2	F	630.50	24.880	4	682.00	43.166	4
	M	626.58	27.966	12	652.00	39.564	12
	Total	627.56	26.465	16	659.50	41.238	16
C4/T1	F	617.67	17.374	6	650.00	22.190	6
	M	613.43	22.090	7	634.29	29.725	7
	Total	615.38	19.354	13	641.54	26.710	13
C4/T2	F	623.83	16.893	6	629.67	20.906	6
	M	604.33	18.556	3	606.00	26.211	3
	Total	617.33	18.960	9	621.78	24.186	9
C4/T3	F	612.60	12.818	5	625.40	29.796	5
	M	610.25	19.328	4	618.50	28.408	4
	Total	611.56	14.959	9	622.33	27.564	9

Table 17 (continued)

School/ Teacher	Gender	Pretest			Posttest		
		Mean	SD	*N*	Mean	SD	*N*
Control Total	F	622.16	22.551	75	655.11	30.533	75
	M	621.50	27.208	64	645.53	36.110	64
	Total	621.86	24.714	139	650.70	33.439	139
E1 /T1	F	640.75	46.126	4	670.75	38.187	4
	M	633.67	24.274	9	651.56	19.944	9
	Total	635.85	30.599	13	657.46	26.735	13
E1/T2	F	630.17	38.207	6	647.83	44.292	6
	M	625.67	51.082	3	656.33	41.885	3
	Total	628.67	39.620	9	650.67	41.021	9
E2/T1	F	591.60	10.262	5	626.00	24.495	5
	M	593.43	29.240	7	631.71	25.824	7
	Total	592.67	22.484	12	629.33	24.302	12
E2/T2	F	610.40	25.909	5	658.60	31.777	5
	M	608.36	25.319	11	641.91	43.924	11
	Total	609.00	24.644	16	647.13	40.241	16
E2/T3	F	593.43	14.853	7	635.14	16.847	7
	M	607.90	33.769	10	637.60	32.864	10
	Total	601.94	27.894	17	636.59	26.749	17
E3/T1	F	605.38	13.763	8	632.88	22.605	8
	M	609.22	24.304	9	632.67	42.226	9
	Total	607.41	19.548	17	632.76	33.393	17
E3/T2	F	601.38	18.662	8	633.75	23.144	8
	M	604.60	16.474	10	627.20	27.680	10
	Total	603.17	17.023	18	630.11	25.247	18
E3/T3	F	610.63	18.283	8	634.63	23.579	8
	M	607.09	17.236	11	630.18	18.930	11
	Total	608.58	17.270	19	632.05	20.503	19

Table 17 (continued)

School/ Teacher	Gender	Pretest Mean	Pretest *SD*	Pretest *N*	Posttest Mean	Posttest *SD*	Posttest *N*
E4/T1	F	638.88	15.029	6	648.50	19.108	6
	M	616.10	16.162	10	628.30	12.074	10
	Total	624.44	18.857	16	635.87	17.640	16
E4/T2	F	629.20	20.606	10	651.90	25.813	10
	M	629.00	27.156	8	647.50	24.646	8
	Total	629.11	22.988	18	649.94	24.656	18
Exp Total	F	614.46	26.669	67	642.67	27.720	67
	M	612.56	26.313	88	636.97	29.708	88
	Total	613.38	26.398	155	639.43	28.913	155
Grand Total	F	618.53	24.793	142	649.24	29.795	142
	M	616.32	26.970	152	640.57	32.719	152
	Total	617.39	25.921	294	644.76	31.587	294

Note: Sch/Teacher identifies teachers with their respective schools (C1= Control school 1).

Table 18

5th Grade Analysis of Covariance: Teachers and Gender

Source	*SS*	*df*	*MS*	*F*	*p*	Partial η^2
Covariate (Pretest)	148397.269	1	148397.269	464.113	.000	.645
Teacher (T)	23870.077	18	1326.115	4.417	.000	.226
Gender (G)	3129.021	1	3129.021	9.786	.002	.037
T x G	3736.611	18	207.589	.649	.858	.044
Error	92130.453	277	332.601			

In the third and final ANCOVA analysis on fifth grade data, there were two independent variables: teacher and gender. Originally, the researcher wished to add ELL as a variable, but many cells had smaller than the required number of cases required for the interaction analysis ($n > 4$). Within this current two-way analysis, 2 of the 38 cells contained a small number of cases ($n = 3$), and therefore the results should be interpreted with some degree of caution. Pretest and posttest means, standard deviations, and group and cell sizes are reported in Table 17. The results of the ANCOVA analysis, as displayed in Table 18, show significant main effects for teacher [$F(1,18) = 4.147$, $p < .000$, partial $\eta^2 = .226$], and for gender [$F(1,1) = 9.786$, $p < .002$, partial $\eta^2 = .037$], but no significant interaction effect between the two variables.

In a further examination of the teacher variable, I employed a one-way factorial ANOVA on pretest scores, which showed a pretest difference among the teachers [$F(1,18) = 3.225$, $p < .000$, partial $\eta^2 = .185$], with a moderately large effect size. A test of equality of error variances showed that the two groups were not homogeneous [Levene's $F(37, 256) = 1.657$, $p < .013$]. Tukey's HSD post hoc analysis of the pretest scores yielded 8 significant differences, with 6 of the significant differences involving one teacher. Teacher E2T1 had significantly lower pretest scores (see Table 15) than C1T2 ($MD = 32.98$, $p < .037$), C2T1 ($MD = 33.58$, $p < .029$), C3T1 ($MD = 34.69$, $p < .048$), C3T2 ($MD = 34.90$, $p < .032$), E1T1 ($MD = 43.18$, $p < .003$), and E4T2 ($MD = 36.44$, $p < .013$). The other two significant differences involved teacher E1T1 who, in addition to having higher scores than E2T1, also had significantly higher pretest scores than E2T3 (*MD* = 33.90, $p < .029$), and E3T2 (*MD* = 32.68, $p < .039$).

I also employed a one-way factorial ANOVA and Tukey's post hoc analysis on math gain scores (posttest minus pretest) by teacher. There was a significant difference among teachers on gain scores [ANOVA $F(1,18) = 4.094$, $p < .000$, partial $\eta^2 = .211$], with a large effect size. A test of equality of error variances showed that the two groups were homogeneous [Levene's $F(18, 275) = 1.618$, $p <.055$].

On math gain scores (posttest minus pretest) by teacher, Tukey's HSD post hoc analysis revealed some significant differences. Teacher C3T2's gain scores were significantly lower than those of seven other teachers: C1T1 (*MD* = 30.56, $p < .008$), C1T2 (*MD* = 31.66, $p < .002$), C2T1 (*MD* = 36.31, $p < .000$), C4T2 (*MD* = 27.49, $p <.035$), E2T1 (*MD* = 32.22, $p <$ =.009), E2T2 (*MD* = 33.68, $p < .002$), and E2T3 (*MD* = 30.20, $p < .008$).

Teacher E4T1's gain scores were significantly lower than those of six other teachers: C1T1 (*MD* = 23.56, $p < .016$), C1T2 (*MD* = 24.66, $p < .008$); C2T1 (*MD* = 29.31, $p < .000$); E2T1 (*MD* = 25.23, $p < .035$), E2T2 (*MD* = 26.69, $p < .005$), and E2T3 (*MD* = 23.21, $p < .031$). Finally, teacher C3T3 had gain scores that were significantly less than the scores of two other teachers: C2T1 (*MD* = 29.97, $p < .006$) and E2T2 (*MD* = 27.35, $p < .031$).

Summary

For the fourth grade, no main effects were found for treatment, gender, or ELL. Scores increased in similar amounts for the control and experimental groups. There was a significant main effect for the school and for teacher, after controlling for the significant pretest differences between schools and teachers, respectively, with the use of pretest scores as a covariate. The effect sizes were moderate to moderately large in each case. There were no significant interaction effects among any of the independent variables tested for the fourth grade (some variables were not crossed in this design and therefore there were no interactions, e.g., school and treatment). Finally, there was a significant main effect on posttest scores for the variable teacher, but there was no significant difference for the variable teacher on the pretest. Unfortunately, the teacher and treatment variables were not crossed in this design, so the significant differences by the teacher on the posttest but not on the pretest could not be tested as an interaction effect.

For the fifth grade, no significant main effect was found for treatment, but there were significant main effects for gender and ELL, even after controlling for the significant pretest differences in each case. In addition, a significant interaction effect was found for treatment and ELL. As with the fourth grade, significant differences were found for school and teacher after controlling for significant pretest differences.

CHAPTER 5

SUMMARY, CONCLUSIONS, AND RECOMMENDATIONS

The No Child Left Behind Act of 2001 was intended to improve the overall quality of education in the United States, with particular emphasis on demographic and educationally specific (special education, English Language Learners) groups that have historically been left behind. Increased assessment and accountability measures have been instituted since the passage of the Act; however, math achievement by ELL students continues to be an achievement gap area.

The purpose of this study was to investigate the effects of using weekly curriculum-based measures on elementary math achievement, with particular emphasis on ELL students. A quasi-experimental quantitative research design was employed to address the following research questions:

1. What is the impact of curriculum-based measures on the math achievement of elementary students?
2. What is the impact of curriculum-based measures on the math achievement of ELL students in particular?

Summary of the Findings

To answer the primary research questions, specific hypotheses were developed and tested. Four null hypotheses were developed for groups at each of two grade levels.

The following four null hypotheses were developed for the fourth grade subjects

Null hypothesis 1a: There will be no statistically significant difference in math scores on the *New Mexico Standards Based Assessment* between the experimental and control groups ($p < .05$).

This hypothesis was tested with an analysis of covariance test and is retained. Overall, there was no difference in the control and experimental groups on the posttest that could be attributed to the use of curriculum-based measures. While no treatment effect was found in the overall analysis, some differences were found (below).

Null hypothesis 1b: There will be no statistically significant difference in math scores on the *New Mexico Standards Based Assessment* among English Language Learner groups ($p < .05$).

This hypothesis was tested with an analysis of covariance test and is retained. There were no significant differences between the current, exit, and never ELL groups.

Null hypothesis 1c: There will be no statistically significant difference in math scores on the *New Mexico Standards Based Assessment* among schools ($p < .05$).

This hypothesis was tested with an analysis of covariance test and is rejected. The results of this investigation revealed that there is a statistically significant difference between schools on the posttest after adjusting for the pretest difference, which favored the control schools (effect size is .127). Two of the experimental schools showed significant differences when compared to the other schools. Of particular interest is the fact that both of these schools had significant differences compared to the other experimental schools. More detailed quantitative and qualitative analysis should be done to try and find possible reasons or explanations to this phenomenon. An initial hypothesis would be that teacher aptitude and attitude played significant roles.

> *Null hypothesis 1d:* There will be no statistically significant difference in math scores on the *New Mexico Standards Based Assessment* among individual teachers ($p < .05$).

This hypothesis was tested with an analysis of covariance test and is rejected. The results of this investigation revealed that there is a statistically significant difference between teachers and the difference is greater than the difference for schools (effect size is .186). Most interesting was that although there was a pretest difference in favor of the control schools, there were no pretest differences among teachers. The conclusion is that certain teachers focused on mathematics instruction and helped their students produce significant student achievement gains.

The following four null hypotheses were developed for the fifth grade subjects

> *Null hypothesis 2a:* There will be no statistically significant difference in math scores on the *New Mexico Standards Based Assessment* between the experimental and control groups ($p < .05$).

This hypothesis was tested with an analysis of covariance test and is retained. Overall, there was no difference in the control and experimental groups on the posttest that could be attributed to the use of curriculum-based measures.

However, within the ANCOVA there were significant differences between gender and ELL (more below).

> Null hypothesis 2b: There will be no statistically significant difference in math scores on the *New Mexico Standards Based Assessment* among English Language Learner groups ($p < .05$).

This hypothesis was tested with an analysis of covariance test and is rejected. The results of this investigation indicate that there were significant differences among ELL groups. The significant difference among ELL groups was both interesting and confusing. As seen in Figure 1 (chapter 4), the current ELL experimental group outperformed the current ELL control group. This was one of the original purposes of using curriculum-based measures, to increase the performance of current ELL students and, when narrowly compared for fifth grade students, there is a significant difference in favor of the students who had access to and used weekly CBMs. However, both the exit and never ELL control groups outperformed their experimental counterparts. A possible explanation is that the curriculum-based measures provided enough language-based math practice to support the needs of the current ELL students, whereas the exit and never ELL students were at a more comfortable level on the use of math language.

> Null hypothesis 2c: There will be no statistically significant difference in math scores on the *New Mexico Standards Based Assessment* among schools ($p < .05$).

This hypothesis was tested with an analysis of covariance test and is rejected. The results of this investigation indicate that there are significant differences between schools on the posttest after adjusting for the pretest difference, which favored the control schools (effect size is .149). Two of the experimental schools scored significantly lower on the pretest, but these results should be interpreted with caution because the groups were not homogenous in terms of variance.

> Null hypothesis 2d: There will be no statistically significant difference in math scores on the *New Mexico Standards Based Assessment* among teachers ($p < .05$).

This hypothesis was tested with an analysis of covariance test and is rejected. The results of this investigation indicate a significant difference among teachers (effect size is .211). Although the teacher groups were not homogeneous on the pretest, they were on the posttest and on gain scores. One control teacher and one experimental teacher both produced significantly lower math gain scores than the other teachers.

For both the fourth and fifth grades, the students in the control and experimental groups had similar gains in math achievement. Furthermore, for the fourth grade there were also no significant effects for gender or ELL. However, this was not the case for the fifth grade, where significant differences were found favoring the girls and there were mixed interactions among the ELL groups. With respect to ELL, the students who are classified as *current* ELL (see Chapter 3) had significantly higher posttest scores than did the control *current* ELL students.

This difference did not occur for the two other ELL groups, as control group students labeled as *exit* and *never* ELL outperformed their respective experimental counterparts.

For both grades, there were significant differences among schools. Raudenbush and Willms (1995) defined the school effect found here as a Type B effect, because it is an effect of actual school practice. This is different from what they describe as a Type A effect, effects that include some school practices but also include contextual influences of social and economic factors (Raudenbush & Willms, 1995). However, the school effect was somewhat smaller than the effect of the teacher.

For the fourth grade, there were no differences among teachers on pretest scores, but there were significant differences on the posttest, and thus on gain scores among teachers. The students of the two teachers at experimental school 1 significantly outperformed students in three other classes, and the students of teacher 2 at experimental school 3 outperformed five other classes. For some of the fourth grade experimental teachers, the use of curriculum-based measures had a significant positive effect on the math achievement of their students.

Significant differences were found among teachers of fifth grade students as well. However, no distinct patterns were identified because the students of one teacher from each group (control and experimental) had significantly lower math gain scores than did the students of other teachers.

Conclusions

Research findings in education are not new or unique that show the classroom teacher is an important, if not the most important, factor in student achievement (examples of recent studies include Berends, Goldring, Stein, & Cravens, 2010; Hill, Rowan, & Ball, 2005; Konstantopoulos, 2009; Nye, Kostantopoulos, & Hedges, 2004; Wayne & Youngs, 2003). The results of the present investigation confirm the previous research—that the classroom teacher is important to student achievement. Wayne and Youngs (2003) found that teachers with certification in, degrees related to, and advanced course work in math produce high school students that performed better in math. Hill, Rowan, and Ball (2005) found that teachers' math knowledge was significantly related to student math achievement gains in elementary school, even after controlling for other variables through the use of covariates. Their recommendation is that one way to improve students' math achievement is to improve teachers' math knowledge. Nye, Konstantopoulos, and Hedges (2004) performed a long-term study that included random assignment of students and found significant teacher effects, with even larger effects on math achievement than on reading achievement. Konstantopoulos (2009) reported on teacher effects and found ample evidence that teacher effects are even more pronounced in schools with high percentages of low socio-economic students. Berends et al. (2010) found that the teacher's focus on academic achievement is significantly related to mathematics gains and that innovative measures, when other factors are controlled for, tend not to be associated with gains in achievement. The results of this research support that of

Berends, et al. (2010) and suggest that quite simply it is the teacher and not the program or curriculum that makes the difference in student achievement, when other factors are controlled.

Recommendations

In terms of further/future research in the use of curriculum-based measures with a population of students, it would be important to either have a within-school control/experimental group setting, or even better have individual teachers use curriculum-based measures with some classes and not with others. This would allow for a better analysis of the effectiveness of curriculum-based measures. It would also be advisable to try and have more detailed information regarding individual teachers so that other independent variables could be crossed and analyzed. As stated previously, ethnicity and socio-economic status were not included as variables in the present study because nearly all the students in the sample were of the same ethnicity and background, so assessing the effectiveness of curriculum-based measures on various ethnic groups and students of different status is also a possible undertaking. Since there was a significant difference for the fifth grade current ELL students, it might be worthwhile to assess larger numbers of current ELL students.

Because the results indicated that teachers had a significant effect on math achievement, it could be important to survey the teachers to determine whether there are differences among them in terms of attitudes towards math, amount of course work in math, and possibly amount of planning time and instructional time that is devoted to math in their classrooms. Administrators and parents concerned with math achievement would be advised to converse with teachers to find out about their attitudes toward math and look at teachers' math results from previous years. Parents should also be more concerned with the effectiveness of their child's teacher than with which school the child attends.

In summary, the effects of curriculum-based measures on elementary students' math achievement was mixed at best; however, the investigation reiterated the importance of the individual classroom teacher on student achievement, emphasizing that the teacher is more important than the school, or the curriculum, *per se*.

REFERENCES

Abedi, J. (2008). Classification system for English language learners: Issues and recommendations. *Educational Measurement: Issues and Practice, 27,* 17-31.

Abedi, J., Courtney, M., Leon, S., Kao, J., & Azzam, T. (2006). *English language learners and math achievement: A study of opportunity to learn and language accommodation* (Tech. No. 702). Los Angeles, CA: National Center for Research on Evaluation, Standards, and Student Testing.

Abedi, J., & Lord, C. (2001). The language factor in mathematics tests. *Applied Measurement in Education*, *14,* 219–234.

Adams, T. L. (2003). Reading mathematics: More than words can say. *Reading Teacher, 56*, 786–795.

Amaral, O. M., Garrison, L., & Klentschy, M. (2002). Helping English language learners increase achievement through inquiry-based science instruction. *Bilingual Research Journal, 26,* 213–239.

Bank Street College. (n.d.). *English Language Learners.* Retrieved November 13, 2009, from http://www.bnkst.edu/literacyguide/ell.html

Baron, R., & Kenny, D. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychological, 51,* 1173–1182.

Basurto, I. (1999). Conditions for reading comprehension which facilitated word problems for second language learners. *Reading Improvement*, *36,* 143-148.

Berends, M., Goldring, E., Stein, M., & Cravens, X. (2010). Instructional conditions in charter schools and students' mathematics achievement gains. *American Journal of Education, 116,* 303–335.

Bloom, H., Richburg-Hayes, L., & Black, A. R. (2007). Using covariates to improve precision for studies that randomize schools to evaluate educational interventions. *Educational Evaluation and Policy Analysis, 29,* 30–59.

Bresser, R. (2003). Helping English language learners develop computational fluency. *Teaching Children Mathematics*, *9,* 294–299.

Brown, C. L. (2005). Equity of literacy-based math performance assessments for English language learners. *Bilingual Research Journal*, *29,* 337–363.

Brown, C. L. (2007). Strategies for making social studies texts more comprehensible for English language learners. *The Social Studies, 98*, 185–188.

Bryk, A. S., Lee, V., & Smith, J. (1990). High school organization and its effects on teachers and students: An interpretive summary of the research. In W. Clune & J. White (Eds.), *Choice and control in American education* (Vol. 1, pp. 135–226). Philadelphia, PA: Falmer.

Buchanan, K., & Helman, M. (1997). *Reforming mathematics instruction for ESL literacy students*. Washington, DC: ERIC Digest.

Buck, G., Mast, C., Ehlers, N., & Franklin, E. (2005). Preparing teachers to create a mainstream science classroom conducive to the needs of English language learners: A feminist action research project. *Journal of Research in Science Teaching*, *42*, 1013-1031.

Bunch, G. C., Abram, P. L., Lotan, R. A., & Valdes, G. (2001). Beyond sheltered instruction: Rethinking conditions for academic language development. *TESOL Journal*, *10*, 28–33.

Cardenas, J., Robledo, M., & Waggoner, D. (1988). *The under-education of American youth*. San Antonio, TX: Intercultural Development Research Association.

Carlo, M. S., August, D., McLaughlin, B., Snow, C. E., Dressier, C., & Lippman, D. (2004). Closing the gap: Addressing the vocabulary needs of English language learners in bilingual and mainstream classes. *Reading Research Quarterly*, *39*, 188–215.

Chamot, A., Dale, M., O'Malley, M., & Spanos, G. (1992). Learning and problem solving strategies of ESL students. *Bilingual Research Journal*, *16*, 1–34.

Chang, M., Singh, K., & Fuller, K. (2009). Language factors associated with achievement grouping in math classrooms: A cross-sectional and longitudinal study. *School Effectiveness and School Improvement*, *20*, 27–45.

Christie, K. (2002). States address achievement gaps. *Phi Delta Kappan*, *84*, 102-103.

Cocking, R. R., & Chipman, S. (1988). Conceptual issues related to mathematics achievement of language minority children. In R. Cocking & J. Mestre (Eds.), *Linguistic and cultural influences on learning mathematics* (pp. 17-46). Hillsdale, NJ: Lawrence Erlbaum Associates.

Collins, A. (1992). Toward a design science of education. In E. Scanlon & T. O'Shea (Eds.), *New directions in educational technology.* Berlin: Springer-Verlag.

Collins, A., Joseph, D., Bielaczyc, K. (2004). Theoretical and methodological issues. *The Journal of the Learning Sciences, 13,* 15-42.

Complete K-12 math standards document. (2008, January). *New Mexico Public Education Department.* Retrieved January 29, 2011, from http://www.ped.state.nm.us/MathScience/mathStandards.html

Core/.basal instructional material adoption approval. (2006, July 21). *New Mexico Public Education Department.* Retrieved January 29, 2011, from http://www.ped.state.nm.us/InstructionalMaterial/d108/adopted/AdoptedMultipleList-Math.pdf

Cronbach, L. J. (1982). *Designing evaluations of educational and social programs.* San Francisco, CA: Jossey-Bass.

Cuevas, P., Lee, O., Hart, J., & Deaktor, R. (2005). Improving science inquiry with elementary students of diverse backgrounds. *Journal of Research in Science Teaching*, *42*, 337-357.

Cummins, J. (1984). *Bilingualism and special education: Issues in assessment and pedagogy*. Austin, TX: Pro-Ed.

Datnow, A., Borman, G. D., Stringfield, S., Overman, L. T., & Castellano, M. (2003). Comprehensive school reform in culturally and linguistically diverse contexts: Implementation and outcomes from a four-year study. *Educational Evaluation and Policy Analysis*, *25*, 143-170.

Denton, C. A., Anthony, J. L., Parker, R., & Hasbrouck, J. (2004). Effects of two tutoring programs on the English reading development of Spanish-English bilingual students. *The Elementary School Journal, 104,* 289-305.

Dixon, J. K. (1995). Limited English proficiency and spatial visualization in middle school students' construction of the concepts of reflection and rotation. *Bilingual Research Journal*, *79,* 221-247.

Dong, Y. R. (2004). Don't keep them in the dark! Teaching metaphors to English language learners. *English Journal*, *93*, 29-34.

Echevarria, J., Vogt, M., & Short, D. (2000). *Improving the quality of instruction for English language learners.* The SIOP Institute. Retrieved October 9, 2010, from http://www.siopinstitute.net/about.html

Echevarria, J., Vogt, M., & Short, D. (2007). *Making content comprehensible for English learners: The SIOP model*. Boston, MA: Allyn & Bacon.

Edwards, V. B. (2009, January 8). "50-state report card." *Education Week, 28*(17), 7.

Egbert, J., & Simich-Dudgeon, C. (2007). Providing support for non-native learners of English in the social studies classroom. *The Social Studies*, *92,* 22-25.

Elementary and Secondary Education Act, Pub.L. 89-10, 79 Stat. 27, 20 U.S.C. ch.70 (1965).

Ernst-Slavit, G., Moore, M., & Maloney, C. (2002). Changing lives: Teaching English and literature to ESL students. *Journal of Adolescent and Adult Literacy*, *46*, 118-128.

Fradd, S. H., & Lee, O. (1999). Teachers' roles in promoting science inquiry with students from diverse language backgrounds. *Educational Researcher*, *28,* 14-20.

Fry, R. (2008). *The role of schools in the English language learner achievement gap* (Rep.). Washington, DC: Pew Hispanic Center.

Gersten, R., & Jimenez, R. T. (1994). A delicate balance: Enhancing literature instruction for students of English as a second language. *The Reading Teacher*, *47,* 438-449.

Gibbs, J. (1994). *Tribes: A new way of learning together.* Santa Rosa, CA: Center Source Publications.

Goldenberg, C., & Rutherford-Quach, S. (2010). *The Arizona home language survey: The identification of students for ELL services.* Civil Rights Project/Proyecto Derechos Civiles. Retrieved October 9, 2010, from http://www.civilrightsproject.ucla.edu

Gunn, B., Biglan, A., Smolkowski, K., & Ary, D. (2002). The efficacy of supplemental instruction in decoding skills for Hispanic and non-Hispanic students in early elementary school. *The Journal of Special Education, 34,* 90-103.

Haager, D., & Windmueller, M. P. (2001). Early reading intervention for English language learners at-risk for learning disabilities:Student and teacher outcomes in an urban school. *Learning Disability Quarterly, 24,* 235-250.

Halliday, M. (1978). *Language as social semiotic: The social interpretation of language and meaning*. London: E. Arnold.

Hart, J. L., & Lee, O. (2003). Teachers’ professional development to improve the science and literacy achievement of English language learners. *Bilingual Research Journal*, *27,* 475-501.

Helwig, R., Rozeck-Tedesco, M. A., Tindal, G., Heath, B., & Almond, P. (1999). Reading as access to mathematics problem solving on multiple-choice tests for sixth graders. *The Journal of Educational Research, 93*, 113-125.

Hill, H. C., Rowan, B., & Ball, D. B. (2005). Effects of teachers’ mathematical knowledge for teaching on student achievement. *American Educational Research Journal, 42,* 371-406.

Hoffer, T., Rasinski, K., & Moore, W. (1995). *Social background differences in high school mathematics and science course-taking and achievement* (Rep. No. 95-206). Washington, DC: U.S. Department of Education, Office of Educational Research and Improvement.

Identifying English-language learners. (2009, January 8). *Education Week, 28*(17), p. 31.

Kamps, D., Abbott, M., Greenwood, C., Areaga-Mayer, C., Willis, H., Longstaff, J., et al. (2007). Use of evidence-based, small-group reading instruction for English language learners in elementary grades: Secondary-tier intervention. *Learning Disability Quarterly, 30*, 153-168.

Kim, S., & Chang, M. (2010). Does computer use promote the mathematical proficiency of ELL students? *Journal of Educational Computing Research*, *42*, 285-305.

Kindler, A. (2002). *Survey of the states' limited English proficient students and available educational programs and services, 2000-2001 Summary Report* (Rep.). Washington, DC: U.S. Department of Education.

Konstantopoulos, S. (2009). Effects of teachers on minority and disadvantaged students’ achievement in the early grades. *The Elementary School Journal, 110*, 92-113.

Lager, C. A. (2004). *Unlocking the language of mathematics to ensure our English learners acquire algebra* (Rep. No. PB-006-1004). Los Angeles, CA: University of California.

Lager, C. A. (2006). Types of mathematics-language reading interactions that unnecessarily hinder algebra learning and assessment. *Reading Psychology*, *27*, 165-204.

Laturnau, J. (2003). Standards-based instruction for English language learners. In G. G. Garcia (Ed.), *English learners reaching the highest level of English literacy* (pp. 286-305). Newark, DE: International Reading Association.

Lau v. Nichols, 414 U.S. 563 (1974).

Lemke, J. L. (2003). *MATHEMATICS IN THE MIDDLE*. Retrieved November 03, 2010, from http://academic.brooklyn.cuny.edu/education/jlemke/papers/myrdene.htm

Lessow-Hurley, J. (2000). *The foundations of dual language instruction*. New York: Longman.

Leung, M., Low, R., & Sweller, J. (1997). Learning from equations or words. *Instructional Science, 25*, 37-70.

Lipsey, M. (1988). Practice and malpractice in evaluation research. *Evaluation Practice, 9*, 5-24.

Lipsey, M., & Wilson, D. (1993). The efficacy of psychological, educational, and behavioral treatment: Confirmation from meta-analysis. *American Psychologist, 48*, 1181-1209.

Littlejohn, J. (1998). *Federal control out of control: The office for Civil Rights' hidden policies on bilingual education* (Rep.). Sterling, VA: Center for equal Opportunity. (ERIC Document Reproduction Service No. ED426598)

May, H., Perez-Johnson, I., Haimson, J., Sattar, S., & Gleason, P. (2009). *Using state tests in education experiments: A discussion of the issues* (NCEE 2009-013). Jessup, MD: National Center for Education Evaluation and Regional Assistance Institute of Education Sciences, U.S. Department of Education.

McBride, A. (2008, summer). *Addressing achievement gaps: The language acquisition and educational achievement of English-language learners. ETS policy notes.* (Vol. 16, No. 2). Educational Testing Service.

McDonnell, L. M., Burstein, L., Ormseth, T., Catterall, J. M., & Moody, D. (1990). *Discovering what schools really teach: Designing improved coursework indicators* (Rep. No. R3886). Santa Monica, CA: RAND.

Mestre, J. P. (1988). The role of language comprehension in mathematics and problem solving. In R. R. Cocking & J. P. Mestre (Eds.), *Linguistic and cultural influences on learning mathematics* (pp. 201-220). Hillsdale, NJ: Lawrence Erlbaum Associates.

Moschkovich, J. (1999). Supporting the participation of English language learners in mathematical discussions. *For the Learning of Mathematics*, *19*, 11-19.

Moving toward language proficiency. (2009, January 8). *Education Week*, *28*(17), 32.

National Council of Teachers of English. (2008). *English Language Learners* [Scholarly project]. Retrieved March, 2010, from http://academic.evergreen.edu/curricular/mit2010/handouts/winter/contentlit/ELLPolicyStatement.pdf

New Mexico Standards Based Assessment (SBA). (2010). New Mexico Public Education Department. Retrieved August 10, 2010, from http://www.ped.state.nm.us/AssessmentAccountability/AcademicGrowth/NMSBA.html

New Mexico standards-based assessment technical manual: Spring 2009 administration. (2010). San Antonio, TX: Pearson Psychometric and Research Services.

NMPED school fact sheets. (2010, December, 20). *New Mexico Public Education Department.* Retrieved January 29, 2011, from http://www.ped.state.nm.us/IT/schoolFactSheets.html

Nye, B., Konstantopoulos, S., & Hedges, L.V. (2004). How large are teacher effects? *Educational Evaluation and Policy Analysis, 26,* 237-257.

Oakes, J., Gamoran, A., & Page, R. N. (1992). Curriculum differentiation: Opportunities, outcomes, and meanings. In P. W. Jackson (Ed.), *Handbook of research on curriculum* (pp. 570-608). New York, NY: Macmillan.

O'Halloran, K. L. (2000). Classroom discourse in mathematics: A multisemiotic analysis. *Linguistics and Education*, *10*, 359-388.

Orr, L. (1999). *Social experiments: Evaluating public programs with experimental methods.* Thousand Oaks, CA: Sage.

Passel, J., & Cohn, D. (2008, February). *Immigration to play lead role in future U.S. growth.* Pew Research Center. Retrieved August, 2010, from http://pewresearch.org/pubs/729/united-states-population-projections

Pawson, R., & Tilley, N (1997). *Realistic evaluation.* Thousand Oaks, CA: Sage.

Peregoy, S. F., & Boyle, O. F. (2005). *Reading, writing, and learning in ESL: A resource book for K-12 teachers*. Boston: Pearson/Allyn and Bacon.

Petrosino, A. (2000). Mediators and moderators in the evaluation of programs for children: Current practice and agenda for improvement. *Evaluation Review, 24,* 47-72.

Pimm, D. (1987). *Speaking mathematically: Communication in mathematics classrooms*. London: Routledge.

Pituch, K. A. (2001). Using multilevel modeling in large-scale planned variation educational experiments: Improving understanding of intervention effects. *The Journal of Experimental Education, 69,* 347-372.

Providing language instruction. (2009, January 8). *Education Week*, *28*(17), 25.

Raudenbush, S. W., & Willms, J. D. (1995). The estimation of school effects. *Journal of Educational and Behavioral Statistics, 20,* 307-335.

Rice, D. C., Pappamihiel, N. E., & Lake, V. E. (2004). Lesson adaptations and accommodations: Working with native speakers and English language learners in the same science classroom. *Childhood Education*, *80,* 121-128.

Rock, D. A., & Pollack, J. M. (1995). *Mathematics course-taking and gains in mathematics achievement* (Rep.). Washington, DC: U.S. Department of Education, Office of Educational Research and Improvement, National Center for Education Statistics. (ERIC Document Reproduction Service No. ED386499)

Schleppegrell, M. J. (2007). The linguistic challenges of mathematics teaching and learning: A research review. *Reading & Writing Quarterly*, *23*, 139-159.

Schon, J., Shaftel, J., & Markham, P. (2008). Contemporary issues in the assessment of culturally and linguistically diverse learners. *Journal of Applied School Psychology*, *24*, 163-189.

Seltzer, M. (1994). Studying variation in program success. *Evaluation Review*, *18,* 342-361.

Sfard, A., Nesher, P., Streefland, L., Cobb, P., & Mason, J. (1998). Learning mathematics through conversation: Is it as good as they say? *For the Learning of Mathematics*, *18*, 41-51.

Shanahan, T., & Beck, I. (2006). Effective literacy teaching for English language learners. In D. August & T. Shanahan (Eds), *Developing literacy in second-language learners: Report of the national literacy panel on language-minority children and youth* (pp.415–488). Mahwah, NJ: Lawrence Erlbaum Associates.

Shandish, W., Cook, T., & Campbell, D. (2002). *Experimental and quasi-experimental designs for generalized causal inference.* Boston, MA: Houghton Mifflin.

Short, D. (1994). Expanding middle school horizons: Integrating language, culture, and social studies. *TESOL Quarterly*, *28*, 581-608.

Snow, C. (2006). Cross-cutting themes and future research directions. In D. August & T. Shanahan (Eds), *Developing literacy in second-language learners: Report of the national literacy panel on language-minority children and youth* (pp. 631–651). Mahwah, NJ: Lawrence Erlbaum Associates.

Solorzano, R. W. (2008). High stakes testing: Issues, implications, and remedies for English language learners. *Review of Educational Research*, *78*, 260-329.

Spanos, G., Rhodes, N. C., Dale, T. C., & Crandall, J. (1988). Linguistic features of mathematical problem solving: Insights and applications. In R. R. Cocking & J. P. Mestre (Eds.), *Linguistic and cultural influences on learning mathematics* (pp. 221–240). Hillsdale, NJ: Lawrence Erlbaum Associates.

Staub, F. C., & Reusser, K. (1995). The role of presentational structures in understanding and solving mathematical word problems. In C. Weaver, S. Mannes, & C. Fletcher (Eds.), *Discourse comprehension* (pp. 285–305). Hillsdale, NJ: Lawrence Erlbaum Associates.

Szpara, M. Y., & Ahmad, I. (2007). Supporting English language learners in social studies class. *The Social Studies*, *98*, 189-195.

Taylor, J., Stecher, B., O'Day, J., Naftel, S., & Le Floch, K. C. (2010). *State and local implementation of the "No Child Left Behind Act". volume IX--accountability under "NCLB": Final report.* Washington, DC: U.S. Department of Education.

Tevebaugh, T. N. (1998). Mathematics is not a universal language. *Teaching Children Mathematics*, *5*, 214-216.

Thomas, W. P., & Collier, V. P. (2002, September 1). *A national study of school effectiveness for language minority students' long-term academic achievement.* EScholarship: University of California. Retrieved August, 2010, from http://escholarship.org/uc/item/65j213pt

Vaughn, S., Mathes, P., Linan-Thompson, S., Cirino, P., Carlson, C., Pollard-Durodola, S., et al. (2006). Effectiveness of an English intervention for first-grade English language learners at risk for reading problems. *The Elementary School Journal, 107,* 153–180.

Wang, J., & Goldschmidt, P. (1999). Opportunity to learn, language proficiency and immigrant status: Determinants of middle school students' mathematics achievement and growth. *The Journal of Educational Research*, *93*, 101–111.

Wang, J., & Goldschmidt, P. (2003). Importance of middle school mathematics on high school students' mathematics achievement. *The Journal of Educational Research*, *97*, 3–19.

Wayne, A. T., & Youngs, P. (2003). Teacher characteristics and student achievement gains: A review. *Review of Educational Research*, *73*, 89-122.

Webb, L. D. (2006). *The History of American Education: A Great American Experiment.* Upper Saddle River, NJ: Pearson.

Weisman, E. M., & Hansen, L. E. (2007). Strategies for teaching social studies to English language learners at the elementary level. *The Social Studies*, *98*, 180-184.

Weiss, C. H. (1997). How can theory-based evaluation make greater headway? *Evaluation Review, 21,* 501-524.

Wright, L. (1997). Enhancing ESL reading through reader strategy training. *Prospect*, *12*, 15-28.

Yeh, S. (2000). Building the knowledge base for improving educational and social programs through planned variation evaluations. *The American Journal of Evaluation, 21,* 27-40.

APPENDIX A

FOURTH GRADE MATH MONSTER

4th Math Monster 1

Name: ________________________________ Total Points (10): _________

Question 1 (Standard 2.1.4)

What is the value of this expression?

19 – (7 + 2) – 3

A 31

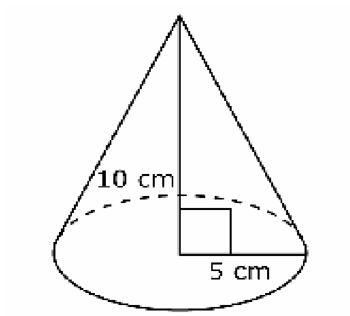
B 13

C 10

D 7

Question 2 (Standard 3.1.1)

A cone is shown in the picture below.



Which of these does not describe the parts of the cone?

A One circular base

B A height of 10 cm

C A diameter of 5 cm

D One vertex

Question 3 (Standard 4.1.2)

Enrique filled a water bottle for his rabbit. Which measure is most likely the amount of water the bottle held?

A 2 pounds

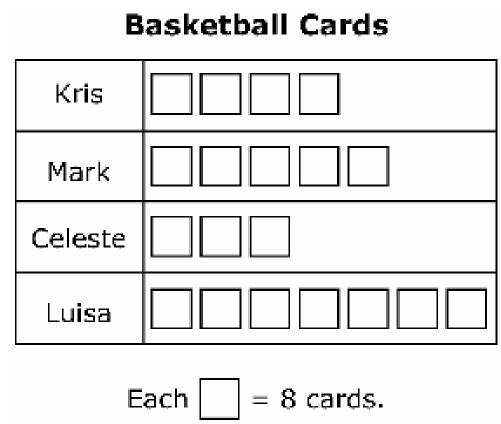
B 2 cups

C 2 gallons

D 2 tablespoons

Question 4 (Standard 5.1.1)

Some of the students in Mrs. Caldwell's class collect basketball cards. The pictograph shows the number of cards each student has.



How many more cards will Kris need to collect to have the same number of cards as Luisa?

A 3

B 8

C 20

D 24

Question 5 (Standard 1.1)

What is the standard form for three hundred six thousand, twenty?

A 300,600,020

B 3,006,200

C 306,200

D 306,020

Question 6 (Standard 1.1)

Which is in the same fact family as this number sentence?

$25 + 75 = 100$

A $100 - 75 = 25$

B $75 - 25 = 50$

C $25 \times 4 = 100$

D $100 + 25 = 125$

Question 7 (Standard 1.3.3) (CR=2pts)

Darius want to spend exactly $20.00 on sports cards. He wants to buy at least one of each card.

Basketball $5.00 per card
Baseball$2.00 per card
Football$3.00 per card

List one combination of sports card Darius can buy using exactly $20.00. Use words, numbers or pictures to explain your answer.

Question 8 (Standard 1.1) (CR=2pts)

Ms. Elliott starts to walk from one end of the Golden Gate to the other end. The total distance is 8,981 feet. After walking 4,765 feet, she stops to rest. How many more feet must Ms. Elliott walk to get to the other end of the bridge?

APPENDIX B

FIFTH GRADE MATH MONSTER

5th Math Monster 1

Name: ______________________________ Total Points (10): _________

Question 1 (Standard 2.1.3)

Alicia wrote this equation.

$y = x + 6$

Which table matches Alicia's equation?

A

x	y
1	7
2	8
3	9

B

x	y
1	6
2	6
3	6

C

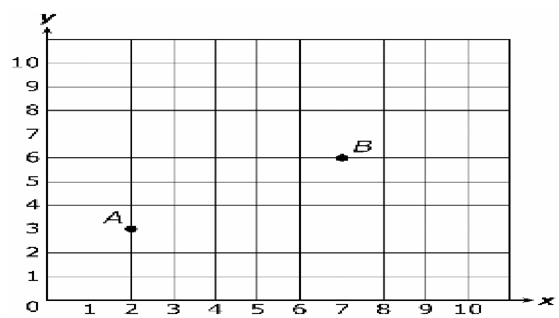
x	y
7	1
8	2
9	3

D

x	y
1	4
2	5
3	6

Question 2 (Standard 3.2)

Thomas plotted points *A* and *B* on the grid shown.



Which ordered pair appears to show the location of Point *A*?

A (2, 3)

B (3, 2)

C (3, 3)

D (7, 6)

Question 3 (Standard 4.2.2)

Which number goes in the ☐ to make the equation true?

$$\begin{array}{r} 76{,}\square 05 \\ -73{,}032 \\ \hline 3{,}573 \end{array}$$

A 7

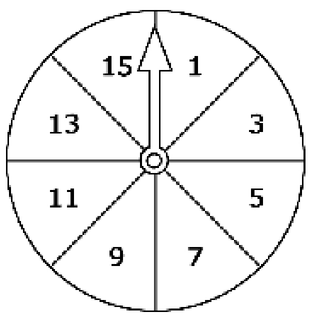
B 6

C 5

D 4

Question 4 (Standard 5.4.2)

All sections of the spinner below are the same size.



What is the probability that the arrow will land on a section marked with a number greater than 10 on the first spin?

A $\frac{13}{15}$

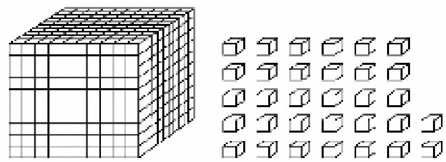
B $\frac{3}{8}$

C $\frac{1}{4}$

D $\frac{1}{8}$

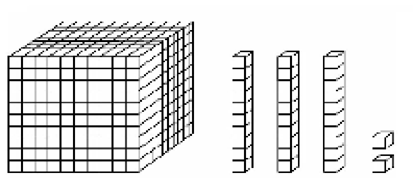
Question 5 (Standard 1.2.5)

Mr. Parker used base-ten blocks to model 1032 for his students.

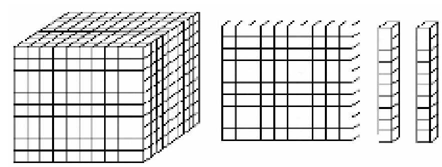


Which of these is another way to model 1032?

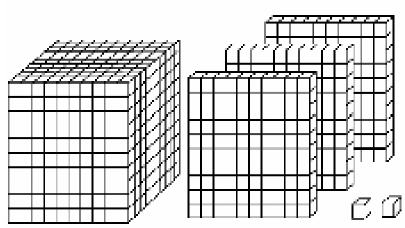
A



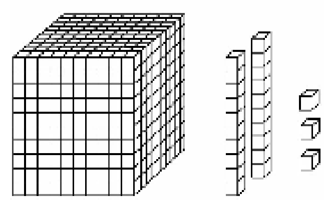
C



B



D



Question 6 (Standard 1.2.5)

Roger saves the same amount of money each week. After 7 weeks, he saved a total of $112. How much does Roger save each week?

A $16

B $17

C $119

D $784

Question 7 (Standard 1.2.5) (CR=2pts)

The manager of a hardware store ordered 200 boxes of nails. Each box contained 50 nails. What was the total number of nails ordered by the manager? Use words, pictures or numbers to explain your answer.

Question 8 (Standard 1.2.5) (CR=2pts)

A pizza maker made one medium pizza and one small pizza for a school party. He used 5 cups of flour for the medium pizza. He used 2 fewer cups of flour for the small pizza than he used for the medium pizza.

Part A. Write an equation that can be used to find the number of cups of flour needed to make one small pizza.

Part B. How many cups of flour are needed to make 7 small pizzas and 9 medium pizzas? Show or explain how you got your answer.

APPENDIX C

INTERNAL REVIEW BOARD LETTER

ASU Knowledge Enterprise Development

Office of Research Integrity and Assurance

To:	Jere Humphreys MUSIC
for **From:**	Mark Roosa, Chair SM Soc Beh IRB
Date:	02/03/2011
Committee Action:	**Exemption Granted**
IRB Action Date:	02/03/2011
IRB Protocol #:	1102005972
Study Title:	The Effects of Curriculum-Based Measures on Elementary Math Achievement

The above-referenced protocol is considered exempt after review by the Institutional Review Board pursuant to Federal regulations, 45 CFR Part 46.101(b)(1) .

This part of the federal regulations requires that the information be recorded by investigators in such a manner that subjects cannot be identified, directly or through identifiers linked to the subjects. It is necessary that the information obtained not be such that if disclosed outside the research, it could reasonably place the subjects at risk of criminal or civil liability, or be damaging to the subjects' financial standing, employability, or reputation.

You should retain a copy of this letter for your records.