PROGRESS MONITORING IN ALGEBRA: EXPLORING RATES OF GROWTH FOR MIDDLE SCHOOL MATH CURRICULUM-BASED MEASUREMENT

by

NANCY J. NELSON WALKER

A DISSERTATION

Presented to the Department of Special Education
and Clinical Sciences
and the Graduate School of the University of Oregon
in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy

June 2011

University of Oregon Graduate School

Confirmation of Approval and Acceptance of Dissertation prepared by:

Nancy Nelson Walker

Title:

"Progress Monitoring in Algebra: Exploring Rates of Growth for Middle School Math Curriculum-Based Measurement"

This dissertation has been accepted and approved in partial fulfillment of the requirements for the Doctor of Philosophy degree in the Department of Special Education and Clinical Sciences by:

Roland Good, Chairperson, Special Education and Clinical Sciences
Elizabeth Harn, Member, Special Education and Clinical Sciences
Leanne Ketterlin Geller, Member, Educational Methodology, Policy, and Leadership
Christopher Phillips, Outside Member, Mathematics

and Kimberly Andrews Espy, Vice President for Research & Innovation, Dean of the Graduate School for the University of Oregon.

June 15, 2011

Original approval signatures are on file with the Graduate School and the University of Oregon Libraries.

An Abstract of the Dissertation of

Nancy J. Nelson Walker for the degree of Doctor of Philosophy

in the Department of Special Education and Clinical Sciences

to be taken June 2011

Title: PROGRESS MONITORING IN ALGEBRA: EXPLORING RATES OF GROWTH FOR MIDDLE SCHOOL MATH CURRICULUM-BASED MEASUREMENT

Approved: ____

Roland H. Good, III, PhD

An examination of evidence-based practices for mathematics reveals that a solid grasp of key algebraic topics is essential for successful transition from concrete to abstract reasoning in mathematics. In addition, experts indicate a need to emphasize formative assessment to allow results to inform instruction. To address the dearth of technically adequate assessments designed to support data based decision making in algebra, this study examined (a) the validity of algebra and mixed computation curriculum-based measurement for predicting mid-year general math and algebra outcomes in 8th grade, (b) growth rates for algebra and mixed computation CBM in the fall of 8th grade, (c) whether slope is a significant predictor of general math and algebra outcomes after controlling for initial skill, and (d) whether growth rates differ for pre-

algebra and algebra students. Participants were 198 eighth grade pre-algebra (*n* = 70) and algebra (*n* = 128) students from three middle schools in the Pacific Northwest. Results indicate moderate relationships between fall performance on mixed computation and algebra CBM and winter SAT-10 and algebra performance and significant growth across the fall. Growth was not found to predict general math and algebra outcomes after controlling for initial skill. Future studies should examine (a) growth rates over an extended period of time with a larger sample of classrooms, (b) instructional variables that may impact growth across classrooms, and (c) the impact on student performance when data gleaned from the mixed computation and algebra CBM are used to support data based decision making in middle school algebra and pre-algebra classrooms.

CURRICULUM VITAE

NAME OF AUTHOR: Nancy J. Nelson Walker

GRADUATE AND UNDERGRADUATE SCHOOLS ATTENDED:

University of Oregon, Eugene, Oregon
San Francisco State University, San Francisco, California
Willamette University, Salem, Oregon

DEGREES AWARDED:

Doctor of Philosophy, School Psychology, 2011, University of Oregon

Master of Arts, Special Education, 2006, San Francisco State University
Level II/Clear Education Specialist Credential, CLAD Certificate

Bachelor of Arts, Psychology, 2002, Willamette University

AREAS OF SPECIAL INTEREST:

Mathematics Instruction and Assessment
Literacy Assessment and Instruction
Response to Intervention
Positive Behavior Support
Early Intervention and Prevention

PROFESSIONAL EXPERIENCE:

Berkeley Unified School District, August 2004 – June 2006
Berkeley High School
Education Specialist & Resource Teacher

Oakland Unified School District, August 2002 – August 2004
Calvin Simmons Middle School
Resource Specialist, Math/Science

GRANTS, AWARDS AND HONORS:

Graduate Teaching Fellowship, 2007-2010
Center on Teaching and Learning, 2009-2010
Teacher Education, Summers 2007 & 2009
Educational Methodology, Policy, and Leadership, 2007-2009
Special Education, 2007-2008

UO Graduate School Research Award, December 2009

College of Education Doctoral Research Award, December 2009
Claire Wilkins Chamberlin Award

College of Education Scholarship, 2009-2010
Dynamic Measurement Group Award

College of Education Scholarship, 2008-2009
Wes Becker Memorial Award

Response to Intervention Training Grant, 2006-2007

PUBLICATIONS:

Harlacher, J. E., Nelson-Walker, N. J., & Sanford, A. (2010). The "I" in RTI: A summary of instructional factors for intensifying interventions. *Teaching Exceptional Children, 42*(6), 30-38.

Nelson-Walker, N. J., Burns, D. A., Turtura, J., Munir-McHill, S. (2010). Dynamic Indicators of Basic Early Literacy Skills: A comparison of Nonsense Word Fluency 6th Edition and Next. Eugene, OR: University of Oregon, College of Education, DIBELS/IDEL Research Team.

Ketterlin-Geller, L. R., Nelson-Walker, N. J., & Jamgochian, E. (2009). Project DIVIDE (Dynamic Instruction via Individually Designed Environments): DIVIDE instructional module development. Eugene, OR: University of Oregon, College of Education, Behavioral Research and Teaching.

ACKNOWLEDGMENTS

I wish to express sincere appreciation to Sisters School District, Springfield Public Schools, and Junction City School District for their participation in this study. I would also like to acknowledge the support I received in preparing this dissertation, from my advisor, Dr. Roland H. Good, III, and Drs. Leanne R. Ketterlin-Geller, and Gina Biancarosa. Special thanks are due to the College of Education and the University of Oregon Graduate School for dissertation awards received, which contributed to the preparation and timely completion of this manuscript.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Table Page

LIST OF GRAPHS

CHAPTER I

INTRODUCTION

According to the 2007 Trends in International Mathematics and Science Study (TIMSS), of 36 and 48 countries that assess students in the 4th and 8th grades, U.S. students ranked 13th and 11th, respectively (NCES, 2009). Based on the results of the same study, only 6% of eighth graders in the United States scored at the *advanced* level in mathematics compared to nearly 40% of the students from the highest performing countries. Within U.S. borders, results from the 2009 National Assessment of Educational Progress (NAEP) indicate that although there are small upward trends in student math performance in grade 8, only 39% and 34% of students scored in the *at* or *above proficient* category in grades 4 and 8 respectively, and gaps between subgroups are growing. For instance, only half as many 8th grade students categorized as socio-economically disadvantaged scored in the *at* or *above proficient* category when compared to peers from higher socioeconomic backgrounds, and only 9% of students with disabilities met the same standard (NAEP, 2009). As evidenced by historical trends of the NAEP and TIMSS assessments, mathematics achievement in the United States is likely to continue to lag behind the achievement of other nations, unless mathematics education in the U.S. receives additional attention in research and practice.

Recommendations for Mathematics Education

In 2001, the National Research Council (NRC) established a panel of experts to confer on the current state of mathematics education in the United States, emphasizing the variability that exists in math standards, curricula, development and foci of assessments, and teaching practices from state to state, ultimately leading to varying degrees of math proficiency between states and other countries. As a result, the NRC made several recommendations, including (a) coordinating curriculum, assessment and instruction materials, and professional development around mathematically-focused school improvement goals; (b) improving students' mathematical learning through coordinated, continual, and cumulative reliance on scientific evidence and systematic evaluation; and (c) conducting additional research "on the nature, development, and assessment of mathematical proficiency" (NRC, 2001, p. 410). In short, the findings of the NRC with respect to the current state of mathematics demonstrate a need for increased understanding about mathematics instruction and assessment.

In congruence with NRC findings, the National Math Advisory Panel (NMAP) released its recommendations in early 2008, identifying specific areas for increased focus in mathematics instruction, highlighting the need to overhaul mathematics curricula, and emphasizing algebra as a point of access for students moving from concrete to abstract mathematics. Recommendations from the final NMAP report demonstrate the need for additional research to support instructional practices in mathematics and a multi-pronged instructional focus to emphasize conceptual understanding, procedural fluency, and automaticity with basic calculations (NMAP, 2008). In addition, the NMAP report identified several areas for continued research, including "ways to enhance teachers' effectiveness, including teacher education, that are directly tied to objective measures of student achievement, and item and test features that improve the assessment of mathematical knowledge" (NMAP, 2008, p. xxvi). In short, the National Math Advisory Panel called for increased study of evaluation procedures used to examine skills that lead to mathematical proficiency.

Other NMAP (2008) recommendations specifically related to assessment include a call for regular use of formative assessment results to design and individualize instruction, national assessments to document and track student achievement and progress on critical topics in math, increased standards for technical adequacy of assessments, and an increased focus on algebra. Consistent with the NMAP report, the National Council of Teachers of Mathematics (NCTM) also recommends that teachers be equipped to identify what students know and need to learn, and supports the use of assessments that provide direct information to teachers and students about the learning process (NCTM, 2000). Based on findings from NRC and NMAP reports, coupled with NCTM recommendations, it is clear that additional research on mathematics assessment designed to advance instruction is critical to increase student math outcomes.

Evidence-Based Practices in Mathematics

Additional expert reports discuss and recommend evidence-based practices for mathematics instruction and assessment (e.g., Gersten, Chard, Jayanthi, Baker, Morphy,

& Flojo, 2008). These practices include the use of explicit, systematic instruction; word problem instruction based on underlying structures; fact fluency instruction; model representations and heuristics; multiple representations and a range of examples; student verbalization of math learning; peer assisted learning; and formative assessment procedures linked to feedback. In addition, experts in the fields of mathematics and education (e.g., Miller & Hudson, 2007; NMAP, 2008; Wu, 1999) indicate instruction must emphasize multiple types of knowledge throughout learning to support adequate understanding of math content, and that precision of language is essential to support student integration of new concepts with already learned material. These practices are described in greater detail in the following paragraphs.

Explicit, systematic instruction. Explicit, systematic instruction is composed of instructional delivery principles and instructional design components (Kame'enui & Simmons, 1990). Instructional delivery principles include using appropriate pacing, allowing adequate processing time, giving students frequent opportunities to respond, monitoring responses, and providiing immediate feedback (Kame'enui & Simmons, 1990). Instructional design principles include targeting big ideas, priming background knowledge, using strategic integration, employing conspicuous strategies, scaffolding learning, and providing judicious review (Kame'enui & Simmons, 1990). Examples of explicit, systematic instruction in the context of mathematics include providing clear models for solving problems using a range of examples; giving students ample opportunities to practice newly learned skills and share mathematical solutions using think alouds; and sharing feedback with students throughout instruction (NMAP, 2008).

Word problem and fact fluency instruction. Word problems provide a unique context for students to grapple with mathematical problems they might face in the real world. In addition, studies that emphasized word problem instruction using underlying structures of word problems found statistically significant increases in student problem-solving skills (Gersten et al., 2009). Fact fluency instruction is also important, because it supports automaticity with basic facts, which reduces the cognitive load required when students attempt to solve word or other complex problems (Ketterlin-Geller, Baker, & Chard, 2008). Although instruction should not focus entirely on fact fluency development, it is recommended that mathematics interventions include a daily dose of fluency instruction and practice (Gersten et al., 2009).

Model representations and heuristics. Model representations and heuristics are enduring methods for portraying problem attributes or solving processes that can be applied across problems. For example, Jitendra and colleagues (e.g., Jitendra, DiPipi, & Perron-Jones, 2002; Jitendra, Hoff, & Beck, 1999) have studied the use of schema-based problem solving, which employs visual representations of problem-solving processes that can be applied to multiple word problem types (e.g., change problems, compare problems). As previously noted with respect to word problem instruction, it is recommended that instruction represent mathematical ideas using visual images and that representations for problem solving focus on underlying problem structures to increase generalizability of skills (Gersten et al., 2009).

Types of knowledge. Research indicates several types of knowledge are important to the development of skills in mathematics (Miller & Hudson, 2007; Wu, 1999). These types of knowledge include declarative (i.e., knowing "that" with automaticity), conceptual (i.e., understanding major concepts, relationships, and connections), procedural (i.e., knowing "how" to complete a task or process), and conditional (i.e., understanding when and why a strategy can be used; discriminating between problems that allow application of strategies and those that don't). To build these types of knowledge in mathematics, strategies that allow for increased depth of knowledge should be employed (Miller & Hudson, 2007).

Multiple representations and examples. To support student development of conditional knowledge, multiple representations and examples should be utilized throughout instruction. Moreover, examples used to illustrate mathematical ideas should be carefully chosen to allow students to develop foundational knowledge before students grapple with non-examples (Kame'enui & Simmons, 1990). In addition, Concrete-Representational-Abstract (CRA) sequences can be used to support depth of mathematical knowledge (Miller & Hudson, 2007). Using concrete strategies (i.e., three-dimensional objects and experiences) builds students' conceptual knowledge. Representational strategies (i.e., two-dimensional pictures, drawings, or diagrams) support students to develop procedural skills. Abstract strategies (i.e., solving problems without pictorial representations or manipulatives) are conducive to ample practice, which supports fluency development.

Precision of language. It is important that precise language is used when teaching students mathematical content because (a) mathematics content is vocabulary-laden, (b) students may commit definitions to memory for later recall, and (c) if slightly inaccurate definitions are used, students may confuse concepts and strategy application. Research indicates that for language to be precise it must be mathematically correct, developmentally appropriate, and longitudinally coherent (Ball & Bass, 2002).

Student verbalizations, peer assisted learning, and motivational strategies. Research also indicates students who use think aloud or question asking strategies for solving problems have improved math performance relative to students who are not taught to use these strategies (Gersten et al., 2008). Similar to recommendations for explicit, systematic instruction, it is suggested that teachers model how to use these strategies using examples and non-examples to support accurate application of these strategies (Gersten et al., 2009). Studies suggest peer assisted learning contributes to increased student outcomes when students are trained to monitor and provide feedback on each other's responses (e.g., Calhoon & Fuchs, 2003). Motivational strategies (e.g., reinforcing student engagement, rewarding students for accomplishments, having students chart their progress) may also support student learning in mathematics, especially for students who may have experienced prior failure or difficulties learning mathematical content (Gersten et al., 2009).

Formative assessment. Formative assessment allows educators to obtain information about student performance during the learning process. Information gleaned from this type of assessment can be used to provide feedback to students about their performance and can support teacher decisions about instructional grouping and student understanding of recently taught content. Research suggests when the results of formative assessment are used by teachers to make instructional modifications and are communicated to students, student performance increases (Gersten et al., 2008). In addition, a recent report from the Institute of Education Sciences (IES; Gersten et al., 2009) indicates two types of formative assessment (i.e., screening and progress monitoring) should be a component of any multi-tiered system of instructional support. Given the federal demand for all students to be proficient in mathematics by 2014 (No Child Left Behind, 2001) and the implications for the role of formative assessment in evaluations for special education service provision (Individuals with Disabilities Education Improvement Act, 2004), school systems are beginning to consider how mathematics fits within a multi-tiered approach. Therefore, a focus on technically adequate formative assessment designed to support instructional decision-making in mathematics is particularly crucial.

Algebra as a Critical Topic

Although considerably less research has examined mathematics education when compared to the area of reading, several major themes have been identified in mathematics that can be connected to a broad set of important math skills (e.g., Gersten, Jordan, & Flojo, 2005; NMAP, 2008). For example, number sense has emerged as an overarching skill in mathematics, foundational to the development of mathematical knowledge, despite a lack of definitional consensus about the construct (Gersten et al., 2005). The literature also reveals some support that students must master basic skills as a means of reducing cognitive load when they grapple with higher-level mathematical problems (e.g., Gersten & Chard, 1999; Ketterlin-Geller et al., 2008; Wu, 1999). Finally, research suggests that algebra serves as the gateway to higher mathematics (Ketterlin-Geller et al., 2008), providing a bridge between concrete and abstract reasoning (Wu, 2009). In addition, students who understand algebra tend to be more prepared for college, higher-level math courses, and jobs in a technology-driven economy (Checkley, 2001). A number of states require a specific course in algebra for high school graduation, 25% of all students take algebra prior to entry to high school, and those students who take a course in algebra prior to high school entry are more likely to enroll in advanced mathematics coursework (Education Commission of the States, 1997). As states move toward requiring algebra coursework and/or proficiency for graduation from high school, adequate preparation for algebra in the early grades and in middle school becomes particularly salient.

Student mathematical proficiency—the fundamental goal of mathematics instruction—is defined as having five independent, yet critical strands, which include (a) conceptual understanding, (b) procedural fluency, (c) strategic competence, (d) adaptive reasoning, and (e) productive disposition (NRC, 2001). The NCTM (2000) reported that students should be able to problem-solve, reason, communicate, connect, and represent concepts in the areas of number and operations, measurement, algebra, geometry, and data analysis and probability; however, these are specific skills within broad categories of mathematics that do not take into account the integrated nature of math learning (Milgram, 2005; Wu, 1999). To initiate a more coherent approach to mathematics instruction, in 2006 the NCTM developed a set of focal points for each grade level targeting increased focus on fewer topics each year (NMAP, 2008); however, it appears that most math curricula have yet to follow suit: At present, mathematics instruction in the United States is largely aligned with textbooks that address topics "a mile wide and an inch deep" (Schmidt et al., 2007).

Perhaps as a result of the surface level content coverage in textbooks and the fragmented nature of instruction in many mathematics classrooms (Schmidt et al., 2007), there is some confusion about the nature of algebra and the topics that comprise a solid initial course in algebra. The NMAP (2008) final report distinguished algebra from school algebra, indicating that although the definition of algebra may be broad, school algebra can be defined more narrowly in the context of Major Topics of School Algebra (i.e., symbols and expressions, linear equations, quadratic equations, functions, algebra of polynomials, and combinatorics and finite probability). In addition, mathematicians indicate that five underlying topics in algebra are essential for effective instruction in mathematics: (a) variables and constants, (b) decomposing and setting up word problems, (c) symbolic manipulations, (d) functions, and (e) inductive reasoning and mathematical induction (Milgram, 2005). These topics in basic algebra provide the foundation for the development of broad mathematical knowledge; thus, students need instruction on these topics to experience success in mathematics. In addition, research indicates that a focus on rational numbers is important to support students in acquiring the abstract reasoning skills they will need for success in algebra (Wu, 2009). For educators looking to support mathematical proficiency and prepare their students for courses in higher-level mathematics, instructional efforts in algebra (and subsequent assessments designed to examine learning) should be aligned with these expert recommendations.

Study Purpose

Based on the premise that the aforementioned key topics (i.e., Milgram, 2005; NMAP, 2008) are those most critical to the development of enduring mathematical knowledge, it is essential that math teachers accurately identify (a) skills matched to each of these key topics, (b) which of these skills students have mastered, and (c) those skills requiring additional instruction. Consequently, the field of mathematics education needs access to formative evaluation tools capable of yielding results that provide educators with information they need to effectively design and adapt instruction to support student progress in skills leading to mastery of key algebra topics. Founded on the premise that formative evaluation is necessary for teachers to monitor student progress toward understanding mathematical concepts and has been indicated as an evidence-based instructional practice, a focus on algebra is critical for success in mathematics, technically adequate formative evaluation tools designed to inform teachers about student learning in algebra are in their infancy, and instruction in middle school has the potential to impact student trajectories in mathematics, the purpose of this study is to evaluate the technical adequacy of several formative evaluation tools in middle school math with a focus on algebra. The following chapter provides additional discussion of the development, uses, and technical adequacy of formative evaluation, and solidifies the rationale for the present study.

CHAPTER II

LITERATURE REVIEW

In the last chapter, statistics were cited supporting the need for increased emphasis on mathematics education in research and practice. In addition, recommendations from experts in the field were discussed, calling for an augmented focus on algebra in instruction and assessment. Formative evaluation was indicated as a means of assessment necessary to support teachers' access to direct information about student learning and an evidence-based practice for improving student outcomes. This chapter describes formative evaluation in greater detail, provides examples of types of formative evaluation and the decisions that can be made from each, and discusses the literature on formative evaluation in mathematics with respect to student outcomes.

Evaluating the Effects of Instruction

To comprehensively approach the design and delivery of instruction, teachers need assessment tools that provide them with information about student learning, before, during, and after the learning process. Specifically, teachers need assessments designed to identify student misconceptions, support adaptations to instruction, and evaluate student mastery of learning objectives when teachers interpret student results. To this end, multiple types of evaluation are useful in schools, including summative and formative evaluation.

Summative evaluation. Summative evaluation, commonly defined as assessment *of* learning, indicates the extent to which a student learned the objectives associated with the content of the test (Stiggins, 2001). Examples of summative evaluations include end-of-year state tests, national assessments, and chapter or unit tests. Typically, results from summative evaluations allow educators to consider student proficiency with content at the end of an instructional period (Ketterlin-Geller et al., 2008). Results from summative evaluation allow educators to consider answers to questions about student mastery of content. For example, an entity interested in accountability information, such as whether or not students have met mathematics standards at the end of a school year, may ask students to complete a test addressing objectives that should have been taught during the year to answer questions about student mastery of content. At the classroom level, a teacher interested in whether or not students mastered the content taught during an instructional unit may give a chapter test to gain insight into the content students learned as a result of instruction. Although summative evaluation is an important tool for schools to use in reporting and analyzing instructional results, it is outside the scope of this study to examine this topic in depth.

Formative evaluation. In contrast to summative evaluation, which occurs after learning, formative evaluation is described as assessment *for* learning (Stiggins, 2001), where student scores serve as indicators of overall understanding in a topic (Lembke & Stecker, 2007), before or during the learning process. Data gleaned from well-designed formative evaluations allow educators to make decisions about student performance while students are still learning a skill (Kelley, Hosp, & Howell, 2008). Several types of formative evaluation exist, including screening assessments (i.e., brief assessments administered at the beginning of an instructional period to identify students in need of additional instructional support), diagnostic tests (i.e., tests that provide information about depth of understanding of specifically identified concepts), and progress monitoring measures (i.e., frequently administered measures that indicate student progress toward objectives over time).

Screening and diagnostic measures provide useful information for educators looking to identify skills needing instruction. Screening measures are typically efficient measures given to a large group of students that allow educators to answer decisions about student knowledge or skill in a broad mathematical domain. For example, a school may choose to administer a screening assessment to all of its students at the beginning of the year to identify those students who may benefit from additional support to meet end-of-year math goals and subsequently provide these students with extra doses of math instruction. In contrast, diagnostic assessments probe depth of skill in a domain and are administered to a smaller group of students to more narrowly define and hone in on skill and knowledge deficits. Results from diagnostic assessments allow an educator to best identify an appropriate start point for instruction. In the aforementioned example for screening assessments, an intervention teacher may choose to administer to any number of her students a diagnostic assessment after screening procedures have occurred, to determine the specific objectives she should target in her instruction. Results from screening and diagnostic assessments can support system-level and instructional grouping decisions by providing information to educators about the types of services and instruction that students may need in order to meet mathematics goals or standards.

A third type of formative evaluation, progress monitoring, is uniquely positioned to inform instructional decisions on an ongoing basis. By definition, "progress monitoring is a scientifically based process used to assess students' academic performance (relative to a target outcome) and evaluate the effectiveness of instruction" (National Center on Progress Monitoring; NCPM, 2010). In order to support these proposed uses (i.e., to allow educators make accurate decisions from the results of progress monitoring measures over time), it is necessary for measures to be robust indicators of important skills, for forms to have roughly equivalent difficulty, and for student performance to be reliable across forms (Francis, Santi, Barr, Fletcher, Varisco, and Foorman, 2008). Because progress monitoring allows educators to make decisions about the effectiveness of instruction while students are learning, progress monitoring is well positioned to support teachers' decisions to adapt instruction based on student performance results. As a consequence of the described cyclical process of using ongoing data to inform instructional changes (see Figure 1), progress monitoring carries the power to improve student outcomes.

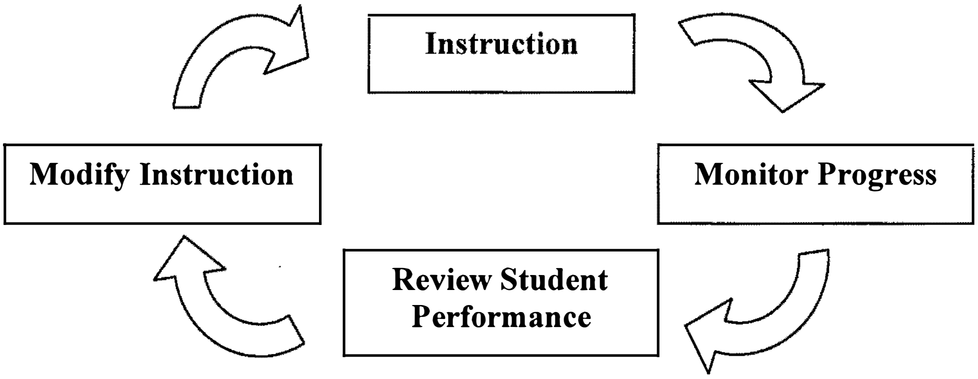


Figure 1. Process for using formative data to inform instructional decisions.

Formative evaluation is essential if educators intend to make timely instructional decisions about the supports needed to ensure students make progress toward annual goals and objectives (Chard, Ketterlin-Geller, Jungjohann, & Baker, 2009). As noted in the last chapter, Gersten et al. (2009) recommends the use of formative assessment for all students in a multi-tiered system of instruction, to identify students at-risk for failure and monitor the effectiveness of interventions for struggling students. In addition, the literature reveals support for formative evaluation as a means for increasing student outcomes in mathematics.

Review of Formative Evaluation Studies

Recent meta-analyses suggest that formative evaluation can have moderate effects in increasing student outcomes when the results of assessments are communicated to teachers and students (Gersten, Baker, & Chard, 2006; Gersten et al., 2008). Fuchs and

Fuchs (1990) studied the use of Curriculum-Based Measurement (CBM) to support instructional decision-making. In the study, teachers in the treatment condition administered CBM probes two times a week for several months. Teachers set goals, graphed student progress using a computer-based program, and received decision rule prompts from the program to adapt instruction or goals given student progress. For those in the basic treatment plus skills analysis condition, teachers were also given data displaying student mastery of skills and problem types attempted. Results revealed strong positive effects for students whose teachers were in the treatment condition containing skills analysis when compared to either the control group or basic treatment condition.

Fuchs, Fuchs, Phillips, Hamlett, and Karns (1995) studied a peer-tutoring program in mathematics that systematically incorporated formative assessment and communicated feedback to teachers and their pupils, resulting in positive outcomes for students. In this study, the teacher assessed student skill weekly using progress monitoring measures. Biweekly, research assistants summarized results of progress monitoring assessments and gave reports to the teacher to support any necessary instructional modifications. Also biweekly, students were given computer-generated graphs of their performance over time and were trained to ask a series of questions about their performance and how they might improve in the next two-week period. During peer tutoring, tutors were trained to give feedback to students whenever they wrote a digit, praising correct responses and providing additional help when tutees responded incorrectly. Combined with other components of the intervention, Fuchs et al. (1995) found positive effects of the peer tutoring program for students when compared to students in the contrast condition.

Calhoon and Fuchs (2003) studied the effects of Peer-Assisted Learning Strategies (PALS) and the use of CBM on math outcomes for high school students with disabilities. In the treatment condition, students completed weekly CBM probes and participated in PALS sessions with peers two times a week for 30 minutes each session. Teachers were given weekly graphs of student performance and class reports summarizing individual student mastery of skills and normative data. The researchers found an effect size of .40 for the PALS + CBM condition using a standardized measure of math computation skills; however, they also found a smaller negative effect for PALS + CBM with respect to student performance on a state high school graduation exam with a greater focus on math application items.

Allinder, Bolling, Oats, and Gagnon (2000) studied teachers uses of curriculum-based measurement across three conditions: (a) control, (b) CBM only, and (c) CBM plus self-monitoring. All students in the CBM conditions were trained to use a computer-based progress monitoring system and taught a test-taking strategy that promoted efficient completion of problems addressing mastered skills. Teachers were taught a set of decision rules for use when examining student performance graphs related to changing student goals and modifying instruction. Teachers in the CBM plus self-monitoring condition had regular conference sessions with research assistants, where they reflected on student performance and developed instructional plans for the next two-week period.

Results of the study indicated that students of teachers in the CBM plus self-monitoring condition made the greatest gains of the three groups on a standardized math computation test, which suggests teacher use of formative assessment data has the potential to positively impact student performance in math.

Fuchs, Fuchs, and Prentice (2004) studied the effects of teaching for transfer and using self-regulation strategies in math problem-solving lessons delivered over three weeks for students with math and reading disabilities. Students were taught rules to reach problem solutions and features of problems that transferred to other problems. Throughout lessons, students scored independent work, charted their scores, set a goal for the next day's work, and were asked to make connections between transfer lessons and real world experiences. The researchers found that students at risk for math and reading disabilities demonstrated less improvement on an outcome measure than did students at risk for math or reading disabilities or no disabilities at all. However, results also indicated that students in the treatment condition made greater improvement than students in the control condition, across all subgroups. Study findings, coupled with authors' previous work examining the same treatment condition (e.g., Fuchs et al., 2003), indicate self-regulation (i.e., feedback to students) is effective for increasing student math outcomes (Fuchs et al., 2004).

Taken together, these studies indicate that formative evaluation plays an important role in mathematics instruction and has the potential to increase student outcomes when results are communicated to teachers and students. Given the critical nature of algebra skill development and the utility of formative evaluation for improving student outcomes, it is important that research examine formative evaluation in the context of algebra, to support the aim of providing teachers with tools that will allow them to be more effective instructors of algebra content. In addition, because progress monitoring has been indicated as a means of formative evaluation that permits timely instructional decision-making and evaluation of annual goals and objectives (Chard et al., 2009), progress monitoring tools for mathematics should be examined.

Progress Monitoring in Mathematics

As defined at the beginning of this chapter, progress monitoring is a method of formative evaluation that is designed to support educators' instructional decisions. Specifically, the results from progress monitoring measures can be used to assess student progress and guide instructional adaptations that support student success. A recommended means for assessing sufficient progress and using performance data to inform instructional decisions is the use of general outcomes measures (GOMs), such as CBM (Burns, Deno, & Jimerson, 2007; Hintze, 2008).

Curriculum based measurement. CBM was developed in the late 1980's as a subset of curriculum-based assessment, with a distinct set of qualities designed to support educators in collecting regular data, graphing student progress, and evaluating student learning over short periods of time (Shinn & Bamonto, 1998). The essential features that distinguish CBM from other formative evaluation methods include its: (a) sensitivity to differences in performance among individuals, (b) sensitivity to differences in performance within individuals over time, (c) use as an indicator of student performance, and (d) focus on basic skills (Shinn & Bamonto, 1998). To support these principles, CBM tasks are typically brief, focus on long-term goals, and utilize standardized administration procedures (Foegen, 2006). Because of the characteristics that make CBM unique, CBM measures can indicate student skill in a topic at a specific time and provide data on how proficiency changes over time (Foegen, 2006). Although originally established to support special education implementation (Fuchs & Shinn, 1989), many teachers use CBM as a method of formative evaluation in reading and mathematics to determine whether instruction and intervention contribute to student learning (Deno, 2002; Deno, Fuchs, Marston, & Shin, 2001; Shinn & Bamonto, 1998).

CBM and mathematics. In mathematics, the bulk of the research examining the use of CBM for progress monitoring has focused on implementation in the elementary grades. Clarke and Shinn (2004) studied several CBM measures designed for early intervention in mathematics and determined that the early math measures were reliable and valid for predicting end of year outcomes in mathematics for first-grade students. Fuchs, Fuchs, Hamlett, Walz, and Germann (1993) examined growth rates for math CBM to support instructional planning in elementary grades. Similarly, VanDerHeyden and Burns (2005) found that the use of CBM for instructional planning resulted in improved student skill in one elementary school. Burns, VanDerHeyden and Jiban (2006) studied the relationship between student performance on the SAT-9 and correct digits on basic and mixed skill probes in an attempt to define categories of performance for math CBM.

In this study, fluency scores demonstrated a moderate positive correlation (r = .55) with SAT-9 performance for 2nd-5th grade students (Burns et al., 2006). In addition, a more recent meta-analysis indicated an overall small to moderate effect size for the use of CBM when the results of progress monitoring assessments were used to inform instructional decisions (Gersten et al., 2008).

Although research on progress monitoring in mathematics is beginning to receive increased attention in the elementary grades, research on the use of CBM to assess secondary mathematics skills is limited. Gersten and Chard (1999) discussed the correlation between number sense and outcomes for secondary students, emphasizing automaticity as a critical skill for success in advanced mathematics. Helwig, Anderson, and Tindal (2002) studied the usefulness of concept-based CBM and found strong correlations between student performance and a computer-adapted outcome measure for students with learning disabilities (r = .61) and general education students (r = .80). More recently, Ketterlin-Geller et al. (2008) have argued for the relevance of CBM performance as an indicator of success in broad math content at the middle school level. Foegen and others (e.g., Foegen, 2000; Foegen & Deno, 2001; Foegen, Jiban, & Deno, 2007) have examined the technical adequacy of CBM measures at the middle school level from a variety of perspectives. However, findings reveal that experts have not reached consensus on the development and foci of math CBMs (e.g., mixed computation algorithms, pre-algebra algorithms, problem solving applications, multiple choice or matching tasks), leaving practitioners with little evidence on which to base instructional decisions (Foegen et al., 2007). Moreover, unlike CBMs for reading, few studies have examined the tenability of CBM for progress monitoring in middle school math.

Studies of rate of change. To determine whether a curriculum-based measure is technically adequate for progress monitoring, three research stages are needed (Fuchs, 2004). These research stages include studies of (a) the technical features of the measure at one point in time (i.e., static performance); (b) the technical features of the rate of change, or slope, of performance; and (c) the utility of the measure for applied use (Fuchs, 2004). To date, much of the research in mathematics CBM has targeted the first research stage; that is, the bulk of math CBM research has explored the validity of measures (e.g., for predicting outcomes on a standardized, norm-referenced test; or with respect to other test items on the measure) as a singular examination of the technical adequacy of the test. Those studies that have occurred at stage two have largely focused on measures appropriate for use with earlier grades.

For example, Hojnoski, Silberglitt, and Floyd (2009) administered four Pre-School Numeracy Indicators (PNIs) monthly in seven Head Start classrooms, between October and May. Hojnoski et al. (2009) found that measures studied demonstrated significant linear growth across the school year, though rates of progress were not examined with reference to an important outcome. Fuchs et al. (1993) also studied math CBMs for evidence of expected slopes. Measures were administered to students in grades one through six weekly in year one and biweekly in year two. The authors suggest, based on significant results for slope, students' scores are expected to grow between .30 and .75 correct digits each week for grades one through six; however, ambitious growth rates of .50 to 1.20 may also be used for goal setting. Thurber, Shinn, and Smolkowski (2002) conducted a confirmatory factor analysis to examine the constructs tested by Mathematics Computation Curriculum Based Measurement (M-CBM). In this study, the researchers identified Computation and Applications as two distinct, yet highly related, constructs assessed by the measures studied, with M-CBM providing moderate evidence of computational skill for fourth grade students (Thurber et al., 2002). Interestingly, Thurber et al. (2002) also found a strong relationship between Reading skill and both Computation and Applications. Clarke and Shinn (2004) studied the concurrent and predictive validity of four early mathematics measures with respect to three criterion outcomes. The researchers collected first grade student performance data for each early math measure at three points in the school year (i.e., fall, winter, and spring) and found significant, positive changes in performance scores across the school year; however, formal growth modeling was not an element of the study design.

Similarly, Foegen (2008) examined the technical features of six middle school math CBMs with respect to end of year outcomes on the Iowa Test of Basic Skills (ITBS). Results of this study indicate increases in student performance across two to three points during the year for each of the measures (i.e., Monitoring Basic Skills Progress—Computation, Concepts and Applications; Basic Facts; Estimation; Complex Quantity Discrimination; and Missing Number) and grades studied (i.e., 6, 7, and 8; Foegen, 2008). Because the study collected data at only three points during the year, included students who missed one administration point, and a considerable amount of variability could be identified in student scores from one data point to the next, additional research is needed to assess whether the slope of progress for any of the measures studied is stable enough to allow for instructional decision-making across the school year.

To move the field of math CBM forward, research should first begin to examine rates of change with respect to measures that provide some evidence for meeting technical adequacy standards in research stage one. In addition, as previously mentioned, the bulk of math CBM research has targeted explorations in the elementary grades. Because middle school math and algebra provide the foundation for later mathematics success, research efforts should prioritize examinations of evaluation and instruction for this age group.

Technical Adequacy of Math Progress Monitoring

Requirements of validity evidence. Messick (1986) identified four facets of test validity: (a) construct validity, (b) values implications, (c) relevance and utility, and (d) social consequences. Good and Jefferson (1998) described these facets in the context of CBM, observing that measures of student performance should not only measure the construct they intend to measure, they should also be directly linked to a purpose and have high benefit relative to the impact on the academic environment. Moreover, Messick (1989) defined validity as "an integrated evaluative judgment of the degree to which empirical evidence and theoretical rationales support the adequacy and appropriateness of inferences and actions based on test scores and other modes of assessment" (Messick, 1989, p. 5). These theories provide the foundation for the imperative that requires an examination of evidence before deeming a test worthy of use.

More recently, the American Educational Research Association (AERA), the American Psychological Association (APA), and the National Council on Measurement in Education (NCME) developed a set of test standards for psychologists and educators, which indicate there are several types of test evidence that may be used to make validity inferences: (a) evidence based on test content, (b) evidence based on response processes, (c) evidence based on internal structure, (d) evidence based on relations to other variables, and (e) evidence based on consequences of testing (AERA, APA, & NCME, 1999). Considering contemporary perspectives of validity in conjunction with documented validity for the general use of CBM (e.g., Deno, 2002; Good & Jefferson, 1998), research should target the validity of CBM for progress monitoring with respect to modern validity standards. That is, studies of the technical adequacy of CBMs must frame research questions and measure features in the context for which the measure will be used, because validity is not attributed to tests or test scores, but to their uses (Kane, 1992).

Necessary features of measures designed for progress monitoring. Because progress monitoring measures are intended to provide information to educators about student changes in performance over time, the technical features associated with progress monitoring may vary with respect to other modes of formative evaluation. For example, while issues of reliability and validity are important for diagnostic and screening assessments, the nature of progress monitoring is such that it requires stringent interpretation of technical features of the test with respect to slope. Francis et al. (2008) define five essential characteristics of progress monitoring assessments: (a) they are administered across regular intervals, (b) they are brief and easy to administer, with moderately little training required, (c) scores should use a consistent metric for interpretability and comparison of scores across test sessions, (d) scores should be predictive of important end of year outcomes, and (e) forms should be free of measurement artifacts so slopes of progress can be attributed to student skills, rather than changes in forms.

In addition, the NCPM (2007) has developed a reference list of criteria for identifying appropriate progress monitoring tools. For progress monitoring assessments to be appropriate, they should be reliable, valid, sensitive to student improvement, linked to improving student learning or teacher planning, have adequate yearly progress benchmarks, and have specified rates of improvement. Of the 34 measures the NCPM reviewed, only six address mathematics skill, and three of these meet the criteria for use with middle school students (NCPM, 2007). Foegen (2006, 2008) recently developed middle school mathematics measures designed to assess general math and algebra skills; however, these measures have not yet been reviewed by the NCPM. To support the use of assessments that are valid for instructional decision-making in mathematics, a more comprehensive analysis of middle school progress monitoring tools is needed to determine whether they are technically adequate for their purported purposes in accordance with NCPM and other validity standards.

The present research will expand on Foegen's work (2000-2009) examining the technical adequacy and associated uses of CBMs for middle school mathematics by studying validity evidence of several math CBMs for use as progress monitoring assessments in 8^{th} grade. To contribute to the research base with respect to the technical adequacy of middle school math CBMs, this study will explore rates of progress for mixed computation and algebra CBMs and examine whether students' slopes can be used to predict their performance on a mid-year math outcome. In the context of modern validity standards, this study aims to examine whether mixed computation and algebra CBMs meet the standards for use as progress monitoring instruments, using evidence based on relations to other variables.

Summary

Because middle school math prepares students for algebra and algebra is considered the gateway to graduation from high school and for other life opportunities (Ketterlin-Geller et al., 2008; NMAP, 2008), it is important that research focus on the development and utility of technically adequate formative measures of middle school math and algebra skills. In addition, to determine whether measures can be used to accurately gauge student progress, studies are needed to explore the sensitivity and anticipated rates of progress for progress monitoring measures in middle school mathematics (Calhoon, 2008). In order to contribute to the research base with respect to current evaluation tools available for use in classrooms for guiding math instruction, this study investigates the validity of middle school math CBM for predicting mid-year mathematics outcomes on a standardized, published, norm-referenced test (See Figure 2). Finally, this study examines expected rates of progress for 8th grade pre-algebra and algebra students on several middle school math CBMs, following the premise that (a) students do not develop abstract reasoning and algebraic skills overnight (i.e., it is expected that pre-algebra students will make more growth on measures of early algebra skill during the course of this study, while students in algebra will make more growth on measures of more advanced algebra skill), and (b) differences in growth rates have important instructional significance. Specifically, this study aims to answer the following research questions:

1. What is the relationship between mixed computation CBM and general math performance? What is the relationship between mixed computation CBM and algebra performance?
2. What is the relationship between algebra CBM and general math performance? What is the relationship between algebra CBM and algebra performance?
3. How much growth can be expected during fall of 8th grade on mixed computation and algebra CBM?
4. Do growth rates predict algebra or general math outcomes, above and beyond initial skills?

5. Are there differences in measure growth rates for pre-algebra or algebra students? Are there differences for the predictive relation between growth rates and algebra or general math outcomes for either group?

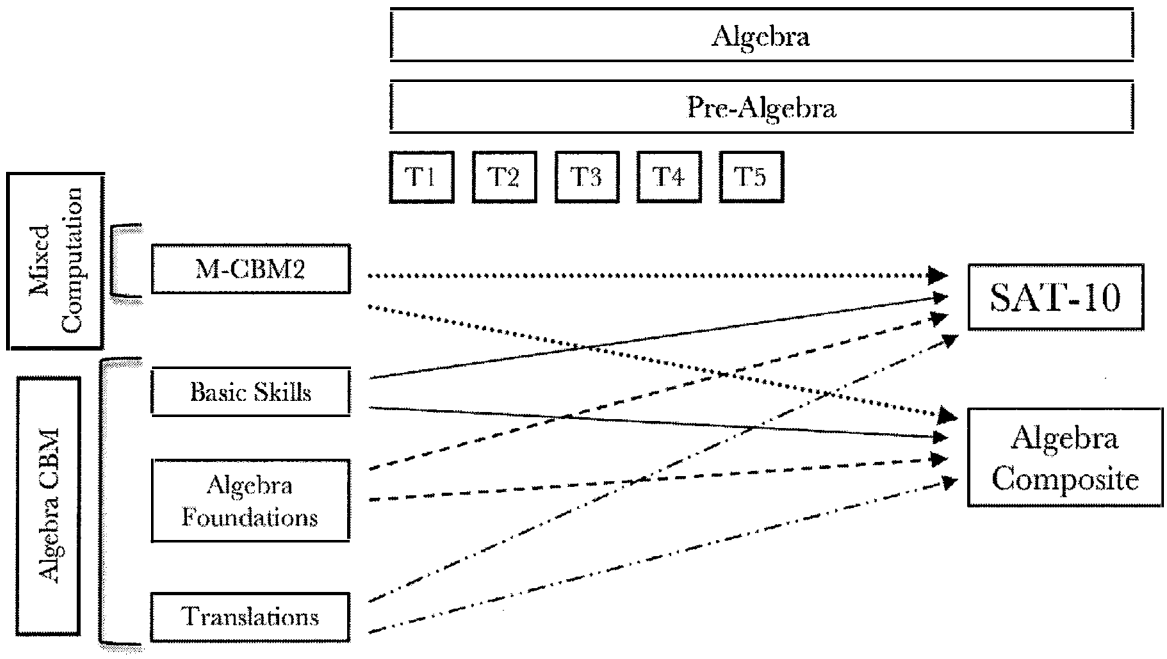


Figure 2. Relationships explored between mixed computation and algebra CBMs and outcome measures for algebra and pre-algebra students in the present study. Mixed Computation represents mixed computation CBM; Basic Skills, Algebra Foundations, and Translations represent algebra CBMs. T1, T2, T3, T4, and T5 represent the measurement occasions that occurred for each of the CBMs. SAT-10 and Algebra Composite represent the two outcome measures administered in the study.

CHAPTER III

METHODS

The first and second chapters of this dissertation reviewed relevant literature and provided a rationale for the present study, which examines rates of progress on mixed computation and algebra CBMs and relations between student performance on these CBMs and math outcome measures for 8^{th} grade students in pre-algebra and algebra classrooms. The third chapter of this manuscript will describe study participants, provide evidence for CBMs administered, describe features of math outcome measures, and elucidate study procedures. A data collection timeline and data analysis procedures will also be described.

Participants

Participants were recruited from three school districts in the northwest region of the United States. All participants were 8^{th} grade pre-algebra or algebra students. After receiving approval from school and district administrators to conduct the study, the researcher obtained recommendations for 8^{th} grade math teacher contacts at middle schools in each district. The researcher contacted teachers via email, explained the study, and requested responses from interested teachers. After the researcher talked with teachers who expressed interest in the study, answered teacher questions about the study, and confirmed that interested teachers met study specifications (i.e., taught 8^{th} grade pre-algebra or algebra courses), the researcher scheduled a visit to explain the study to students, distribute a passive consent letter for parents, and request student assent to participate. In total, three teachers in three schools in three districts (i.e., one teacher per school per district) agreed to allow their students to participate in the study. Across the three schools, ten classrooms met the criteria for 8th grade pre-algebra or algebra content and 232 students and their parents gave consent to participate in the study. Students in five classrooms received pre-algebra instruction; students in the remaining five classrooms received algebra instruction. Demographic data for participants are provided as a function of publicly available data for each school.

School A. School A is the primary middle school for a district located in a small town (i.e., approximately 2,000 people), situated outside a mid-sized city. In 2008-2009, 418 students attended School A. In the same year, 27% of students at School A qualified for free and reduced price lunch, 1% were English Language Learners, and 13% were identified as having disabilities. The student population was categorized as 93% White, 3% Hispanic, and 2.5% Asian/Pacific Islander. The participating teacher at School A was male. Among other courses, he instructed three 55-minute sections of pre-algebra and one 55-minute section of algebra each day. In 2008-2009, School A met the criteria for adequate yearly progress (AYP) for all sub-groups represented at the school.

School B. School B is the primary middle school for a district located in a small city (i.e., approximately 5,000 people), situated outside the second largest metropolitan area in the state. In 2008-2009, 503 students attended School B. In the same year, 51% of students at School B qualified for free and reduced price lunch, 5% were English

Language Learners, and 19% were identified as having disabilities. The student population was categorized as 81% White, 12% Hispanic, 3% American Indian/Alaskan Native, and 1.5% Asian/Pacific Islander. The participating teacher at School B was male. Among other courses, he instructed two 55-minute sections of pre-algebra and two 55-minute sections of algebra each day. In 2008-2009, School B met the criteria for AYP for all sub-groups represented at the school.

School C. School C is one of several middle schools in a district located in a mid-sized city (i.e., approximately 60,000 people), situated in the second largest metropolitan area in the state. In 2008-2009, 617 students attended School C. In the same year, 61% of students at School C qualified for free and reduced price lunch, 7% were English Language Learners, and 16% were identified as having disabilities. The student population was categorized as 72% White, 18% Hispanic, 2.5% Black, 2.5% Multiethnic, 2% American Indian/Alaskan Native, and 2% Asian/Pacific Islander. The participating teacher at School C was male. Among other courses, he instructed two 110-minute blocks of general science and algebra each day. In 2008-2009, School B met the criteria for AYP for all sub-groups represented at the school, except students with disabilities.

Measures

Several measures were used to answer the research questions targeted by this study. The independent variables in the study were student performance on each of the four CBM measures, across measurement occasions: (a) Mathematics Computation Curriculum-Based Measurement, 2nd Version (M-CBM2); (b) Basic Skills; (c) Algebra

Foundations; and (d) Translations. The dependent variable in the study was student performance on the Stanford Achievement Test, 10th Edition, composed of Math Procedures and Math Problem Solving subtests. In addition, an algebra aggregate score was calculated, hereafter referred to as the Algebra Composite, using selected items from the Math Problem Solving subtest of the SAT-10.

M-CBM2. Math Computation Curriculum-Based Measurement, 2nd Version, is a group-administered, four-minute progress monitoring measure, developed by AIMSweb (see Appendix A). AIMSweb provides roughly 30 alternate forms of M-CBM2 for 7th and 8th grades and reports each form is designed to be roughly equivalent in difficulty. The 8th grade M-CBM2 probes examine student skill in addition, subtraction, multiplication, and division of whole numbers, fractions, and decimals; and fraction, decimal, and percent conversions. Standardized directions allow students to skip problems they "really don't know how to do," and tell students to round to the hundredths place and keep fractions in their simplest form (Shinn, 2004). The total score for student performance on M-CBM2 consists of the sum of the number of digits computed correctly in each answer in four minutes. Little information is available from AIMSweb about the technical adequacy of M-CBM2; however, studies of the reliability and validity of M-CBM at grades 3 and 4 demonstrate moderate to strong correlation coefficients. Thurber et al. (2002) found that M-CBM demonstrated strong correlations with measures of basic facts computation (median $r = .82$) and moderate to strong correlations with published, standardized tests ($r = .42$ to $.63$). According to the NCPM (2007), AIMSweb math measures meet the foundational psychometric standards of reliability and validity. Still, M-CBM2 is a relatively new measure, when compared to the M-CBM measure designed for administration to students in grades 1 through 6. Consequently, reliability and validity of M-CBM2 for use as a static indicator and progress monitoring measure in middle school will be further explored in this study.

Basic Skills. The Basic Skills measure (see Appendix B) addresses basic algebra knowledge, by examining student skills in applying proportional reasoning, "solving simple equations, applying the distributive property, working with integers, and combining like terms" (Foegen, Olson, & Perkmen, 2005). Each five-minute, group-administered probe consists of 60 problems, with the total score being the number of problems answered correctly in five minutes (Foegen, 2009). Students receive full credit for responses that are mathematically equivalent to the correct answer. According to the technical manual, in the high school setting, reliability estimates for single form administration of the Basic Skills measure ranges from .71 to .89 (Foegen et al., 2005). When scores are aggregated (i.e., two scores are averaged), the reliability of the Basic Skills measure ranges from .85 to .88. Aggregated scores demonstrated moderate correlations with the student composite scores on the Iowa Algebra Aptitude Test (r = .56) and low correlations with computation subtest scores on the Iowa Test of Educational Development (r = .40). Measure reliability appears to be within the acceptable range for progress monitoring decisions (Salvia & Ysseldyke, 2007); however, validity evidence based on relations to other variables may be insufficient, depending on the content assessed by the criterion measure. Consequently, the validity of the measure will be further examined in this study with respect to standardized measures of general math and algebra achievement for 8th grade students, at one point in time and over the course of the fall of 8th grade.

Algebra Foundations. A five-minute, group-administered measure with 42 items (see Appendix C), Algebra Foundations assesses student skill in four core areas of algebra: (a) reading graphs, (b) evaluating variables and expressions, (c) solving equations and simplifying expressions, and (d) identifying rules from tables and combinations of numbers (e.g., patterns). The total score on the measure is the number of problems answered correctly in five minutes (Foegen, 2009). According to the technical manual, in the high school setting, single form reliability for the Algebra Foundations measure ranges from .59 to .71 (Foegen et al., 2005). Aggregated score reliability for two administrations ranges from .73 to .76. Aggregated scores demonstrate moderate correlations with student composite scores on the Iowa Algebra Aptitude Test (r = .57) and low correlations with computation subtest scores on the Iowa Test of Educational Development (r = .29). Similar to the Basic Skills measure, the strength of validity evidence for Algebra Foundations may depend on the criterion measure. The reliability and validity of Algebra Foundations will be further examined in this study to determine its appropriateness for use in 8th grade classrooms as a static indicator of achievement and progress monitoring measure with respect to general math and algebra outcomes.

Translations. The Translations measure was developed in close alignment with the Connected Mathematics Project (CMP) curriculum, in an effort to examine conceptual applications of algebra with reduced reliance on symbols (Foegen, 2009). The measure is a seven-minute, 42-item assessment of students' skills in matching data tables, graphs, and equations and their knowledge of the effect of slight changes in one component to other relevant information (see Appendix D). To control for guessing on the measure, the total score is computed as the number of problems answered correctly minus the number of problems answered incorrectly in seven minutes. Skipped problems are not counted incorrect. Students are told not to guess on items, as an incorrect guess will penalize their score. According to Foegen et al. (2005), in the high school setting, single form reliability estimates for the Translations measure range from .46 to .66 and aggregated score reliability estimates range from .60 to .72, below the recommended standard for progress monitoring. Aggregated scores demonstrate moderate correlations with student composite scores on the Iowa Algebra Aptitude Test ($r = .35$) and low correlations with the computation subtest of the Iowa Test of Educational Development ($r = .29$); however, these estimates are derived from administration in settings that employ traditional math curricula. Authors of the test speculate that the Translations measure may fare better in a classroom that relies on a concept-based curriculum, such as CMP (Foegen, 2009). In addition, this measure was included in the present study given a theoretical interest in studying assessments of conceptual understanding in mathematics.

Stanford Achievement Test, 10^{th} Edition. The SAT-10 is a standardized, norm-reference test designed to assess academic content knowledge in a range of disciplines from kindergarten through 12^{th} grade. For the purposes of this study, only the 8^{th} grade mathematics subtest was administered. The SAT-10 math subtest is an un-timed, group-administered, multiple-choice assessment, constructed of two subtests: Math Problem Solving and Math Procedures. The math portion of the SAT-10 is expected to take 80 minutes to complete (i.e., 50 minutes for Math Problem Solving, 30 minutes for Math Procedures). Aligned with the NCTM Principles and Standards for School Mathematics, the SAT-10 math subtests measure content and processes in number sense and operations; patterns, relationships, and algebra; geometry and measurement; and data, statistics, and probability (Pearson Education, 2009). Skills in mathematical communication and representation; estimation; mathematical connections; and reasoning and problem solving are also assessed. Scores provided on the SAT-10 include scaled scores, national and local percentile ranks and stanines, grade equivalents, and normal curve equivalents (Pearson Education, 2009); however, only raw scores were used for this study. More information about the development of the SAT-10 can be obtained from Pearson Education.

Algebra Composite. Because it was expected that the SAT-10 would provide a strong indication of student general math performance but might be too diffuse to more narrowly examine content taught in algebra classrooms, an Algebra Composite score was created from items on the SAT-10 to allow for an examination of relations between measure scores and algebra outcomes. To identify the items that would construct the composite score, each Math Problem Solving and Math Procedures item was compared to algebra test items identified by the National Assessment for Educational Progress (NAEP, 2003-2009). In addition, the researcher compared target skills of test items to NCTM Focal Points for grade eight and referenced Milgram's (2005) key topics in algebra.

After reviewing these standards and test items, eight algebra items were identified from the SAT-10 Math Problem Solving subtest to include in the Algebra Composite, because they represent content that is fundamentally algebraic in nature. While other items on the SAT-10 may address algebra skills in combination with other topics, those that were included in the Algebra Composite are unlikely to be disputed as algebra items when considered relative to the other test items and domains purported by SAT-10 publishers. For example, an item requiring students to identify the equation that symbolizes the graphical representation of a linear equation was included in the composite, while an item asking students to identify the perimeter of an object given a figure was not included in the composite. The target skills of each of the items included in the Algebra Composite are listed in Table 1, along with examples of tasks similar to those required by included items. Cronbach's Alpha was computed to examine the internal consistency of items included in the Algebra Composite. The reliability coefficient for the eight Algebra Composite items (α = .68) is approaching the generally accepted value of $\alpha = .70$, which provides some evidence for use of the Algebra Composite as a criterion measure of an algebra-related construct.

Table 1

Target Skills and Example Tasks for SAT-10 Items Included in the Algebra Composite

Item Descriptor	Target Skill	Example Tasks
MPS 21	Extend a pattern involving figures representing perfect squares	Given a visual display of a pattern representing numbers, identify the next two numbers in the sequence
MPS 22	Determine an equation given a table of x and y values	Given a function table for (x, y), identify the equation that represents the ordered pairs in the table
MPS 23	Identify an equivalent algebraic expression	Given an expression containing two unknowns with coefficients, identify an equivalent expression
MPS 24	Solve an algebraic equation (in terms of a variable)	Given an equation containing three terms and one unknown, determine the value of the unknown.
MPS 25	Identify a linear equation from a graph and two ordered pairs	Given a graph of a linear equation and two identified ordered pairs, choose the equation that represents the linear equation
MPS 26	Evaluate an expression for specific values	Identify the value of a ratio expression containing three unknowns when values for the three unknowns are provided
MPS 27	Solve an inequality (in terms of a variable)	Given an inequality containing one unknown, identify a possible value for the unknown

Item Descriptor	Target Skill	Example Tasks
MPS 38	Identify coordinates for a point on a graph of a linear equation	Given a linear equation and its graph, identify an ordered pair the line includes.

Note. MPS = Math Problem Solving. Item descriptor numbers indicate item numbers on the SAT-10. Target skills parallel those defined by NAEP (2003-2009).

Procedures

Prior to beginning the study, the researcher obtained training to administer and score each of the CBM measures used in the study design. To prepare for data collection, the researcher worked with teachers to develop a bi-weekly schedule for CBM and outcome measure administration, with no more than one week between CBM administrations across sites for each data point. Students who missed researcher administration of CBM probes were not permitted to complete missed probes. School staff was not trained to administer each measure, and time and distance constraints did not allow the researcher to travel to each school site a second time each week for additional administration sessions. SAT-10 data collection occurred approximately two months after the final CBM data point at each site. See Table 2 for a data collection timeline for this study.

Table 2

Data Collection Timeline

Data Point	Time	School	Measures Administered*
1	October, week 4	A	M-CBM2
	November, week 1	B and C	Basic Skills
			Algebra Foundations
2	November, week 1	A	Basic Skills
	November, week 2	B and C	Algebra Foundations
			Translations
			M-CBM2
3	November, week 3	All	Algebra Foundations
			M-CBM2
			Basic Skills
4	December, week 1	All	Basic Skills
			M-CBM2
			Algebra Foundations
5	December, week 3	All	M-CBM2
			Basic Skills
			Algebra Foundations
			Translations
SAT-10	February, week 3	A	Math Problem Solving
	February, week 4	B and C	Math Procedures

**Note*. Measures were administered in the order listed at each corresponding administration session.

After a schedule was agreed upon at each school site and consent was obtained from all participants, the researcher prepared packets containing all of the CBM measures to be administered for the duration of the study. Each packet intended for algebra course classrooms contained a cover sheet used to match each student to his/her packet; four forms of three of the CBM measures: (a) M-CBM2, (b) Basic Skills, and (c) Algebra Foundations; and two forms of the Translations measure. For pre-algebra course classrooms, the Translations measure was not included in the packet. The researcher randomized the order of each CBM measure in the packet using a Latin square design (Stewart, 2007) to allow for randomized administration and control of practice effects at each testing session.

Before the first administration, all participants were assigned a random identification number, which was written in the top right corner of each CBM measure and on the packet cover sheet. At the first administration, students were directed to write their name on their cover sheets but not on any of the measures. At each administration, the researcher only removed the CBM measures from each packet that had been completed in that day's testing session. Cover sheets were kept in the classroom in a secure location with packets containing measures not yet completed. Prior to administration of the SAT-10, the researcher recorded the same random identification number used for student packets on SAT-10 teleforms. These were matched to student cover sheets and distributed to students for testing. Cover sheets were stored in the classroom following SAT-10 testing to allow for subsequent data collection permitted by approved IRB # E117-10. Per this IRB, cover sheets will be destroyed before the end of Fall 2010.

During each administration session, the researcher administered the M-CBM2, Basic Skills, and Algebra Foundations measures to all of the students in each classroom. At the 2nd and 5th administrations, the researcher administered the Translations measure to students in algebra classrooms. This measure was only administered twice to account for the number of alternate forms available for the measure while still allowing for an approximation of growth in data analysis. In addition, the Translations measure was only administered to algebra students because of the increased duration of the measure and the difficulty of the tasks included in the measure. Because each of the CBM measures is timed, each testing session lasted 20-30 minutes, depending on the schedule of measures for the testing session. After each administration, the researcher scored completed probes, entered scores into an Excel spreadsheet using random identification numbers, and returned probes and score reports to each school at the subsequent testing session. To facilitate teacher use of student performance data for instructional decisions, the researcher matched student work and score reports to student names using student cover sheets while in each classroom.

In February, the researcher returned to each participating classroom to administer the SAT-10. On the first day of SAT-10 testing, the researcher read standardized administration directions to students and walked through practice examples for the Math Procedures and Math Problem Solving subtests of the SAT-10. For schools A and B, on the first day of SAT-10 testing, students were given at least 50 minutes to complete SAT-10 activities. For schools A and B, on the second day of SAT-10 testing, the teacher allowed students at least 30 additional minutes to complete SAT-10 activities. For school C, students were assessed by the researcher in one day and were given at least 80 minutes to complete the SAT-10. Following the second day of SAT-10 administration for schools A and B and the first day of SAT-10 testing for school C, the researcher collected all SAT-10 booklets and teleforms for scoring and data entry.

According to the SAT-10 manual, it is recommended that students receive at least 80 minutes for completion of the Math Procedures and Math Problem Solving subtests; however, the test is not timed. Although it would have been desirable to allow students as much time as needed to complete all items on the SAT-10, it was not feasible to do so given consent procedures and initial agreements with study sites.

Data Analysis

To answer the first two research questions targeted by this study concerning the strength of the relations between measures, Pearson correlation coefficients were computed and examined for relations between each CBM, SAT-10 subtests, the SAT-10 total score, and the Algebra Composite score. To address research questions three through five for the Translations measure, Multiple Regression was used to examine gain scores between the 2^{nd} and 5^{th} administrations (i.e., those measurement occasions where the Translations measure was administered to students). To answer research questions three through five for the three measures that were administered on more than two measurement occasions (i.e., M-CBM2, Basic Skills, and Algebra Foundations), data were analyzed using the principles of Hierarchical Linear Modeling (HLM) outlined by Raudenbush and Bryk (2002).

HLM allows for simultaneous modeling of predictors in nested samples and more accurate interpretation of variance than is explained by levels of systems. For example, student outcomes following an intervention are likely influenced by the quality of implementation of the intervention, teacher or classroom level variables, and school level variables. To attribute all variance in student outcomes to the intervention does not accurately depict the influence of the system on the outcome. In this study, HLM was used to model student growth over time within several schools, where growth (i.e., performance over time) is expected to vary as a function of individual student learning, which may vary across classrooms. The use of HLM allowed for an examination of growth rates across five measurement occasions for individual students on mixed computation and algebra CBM. In this study, HLM also allowed for an examination of the relationship of student slopes with respect to SAT-10 and algebra outcomes in pre-algebra and algebra classrooms. Specific aspects of HLM analyses (e.g., predictors, modeling decisions, final models) are described in the following chapter.

CHAPTER IV

RESULTS

In chapter three, initial participants, measures of interest, and study procedures were described. Methods for data analysis were outlined, including bivariate correlations, the use of Hierarchical Linear Modeling to study growth rates and slopes of progress relative to SAT-10 and algebra mid-year outcomes, and Multiple Regression to model student performance gains using the Translations measure. In this chapter, participants will be further defined following procedures for managing missing data and descriptive statistics will be provided for each predictor of interest in the study. Analyses and results will be described in the context of study research questions and general implications of results will be discussed.

Missing Data

To prepare data for analysis, missing data were examined. Of the initial sample of 232 students, 219 participated in the SAT-10 assessment. Scores for the 13 students who did not take the SAT-10 were not included in the analysis. Of the remaining 219 students in the sample, 21 students missed more than one full day of CBM testing. A comparison of SAT-10 total score means for students missing more than one full day of CBM testing and the rest of the sample was not significant, $F(2, 217) = 0.01$, $p = .92$. In addition, comparing SAT-10 subtest and total score means by classroom, school, and content did not reveal significant results, suggesting differences in SAT-10 means for students who missed more than one full day of CBM testing were no greater than they would be by chance. Significant differences between students missing more than one full day of CBM testing and the rest of the sample could not be identified with respect to outcome variables; thus, those students who missed more than one full day of CBM testing were removed from the data set, to allow for more accurate modeling of growth across the sample for all measurement occasions. The final sample included 198 pre-algebra (n = 70) and algebra (n = 128) students across ten classrooms in three schools. The median number of participants per classroom was 23 (range = 3 - 32) and the median number of participants per school was 53 (range = 53 - 92).

Descriptive Statistics: Final Sample

CBMs. Descriptive statistics for each measurement occasion of the M-CBM2, Basic Skills, Algebra Foundations, and Translations measures for all participants in the final sample are provided in Table 3. Mean scores across measurement occasions demonstrate a general upward trend, which suggests student math performance improved over time. Overall, standard deviations are acceptable for the sample; however, the size of the standard deviation relative to the mean score for the first administration of the Translations measure indicates moderate variability in student performance on the measure at the first administration. An examination of distributions indicates roughly normal distributions for each measure across measurement occasions; however, some positive skew is associated with the Translations measure, indicating a number of students earned scores of zero on the measure at both data points.

Closer inspection of mean scores indicates some trajectory differences for M-CBM2, Algebra Foundations, and Basic Skills. For the M-CBM2 and Algebra Foundations measures, there is a slight decline in mean scores at the second administration. For the Algebra Foundations measure, the mean score for the third administration is also below the mean score for the first administration. For both measures, if the first administration were to be removed from the data set, remaining mean scores would indicate a positive trend across all administrations. For the Basic Skills measure, there is considerable variability in mean scores across administrations. Between the first and fifth administrations there was a small increase in student scores; however mean scores do not indicate a stable positive trend over time. See Graph 1 for a visual display of mean scores across data points.

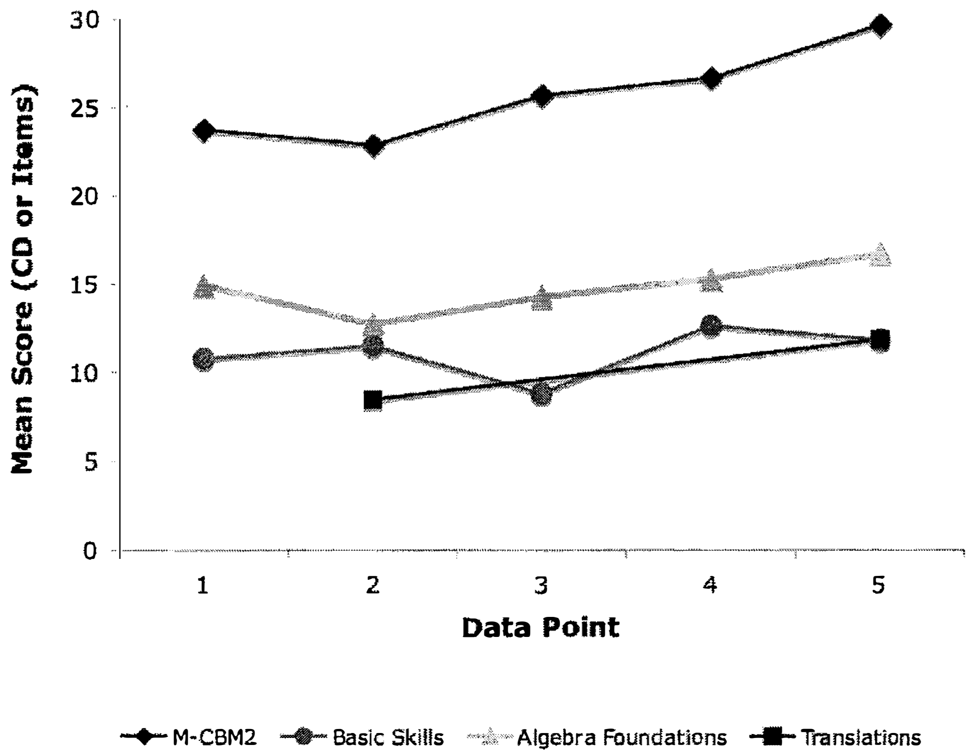
Table 3

Descriptive Statistics for Algebra and Mixed Computation CBM by Time

Measure/Time	*n*	*Min*	*Max*	*M*	*SD*
M-CBM2					
Time 1	187	0	61	23.74	10.01
Time 2	187	3	66	22.87	9.12
Time 3	193	0	68	25.69	11.44
Time 4	194	0	74	26.70	12.27
Time 5	184	1	87	29.70	13.53

Measure/Time	*n*	*Min*	*Max*	*M*	*SD*
Basic Skills					
Time 1	187	0	26	10.80	5.63
Time 2	187	1	32	11.52	5.78
Time 3	193	0	27	8.81	5.83
Time 4	195	0	37	12.66	6.56
Time 5	184	1	30	11.84	6.10
Algebra Foundations					
Time 1	187	0	32	15.01	6.23
Time 2	187	0	33	12.82	5.82
Time 3	193	0	31	14.36	6.52
Time 4	194	0	35	15.37	6.98
Time 5	184	3	40	16.79	7.98
Translations					
Time 2	120	0	27	8.49	7.40
Time 5	120	0	32	11.88	7.48

Outcome measures. Descriptive statistics for the SAT-10 subtests (i.e., Math Procedures and Math Problem Solving), SAT-10 total score, and Algebra Composite are provided in Table 4. Standard deviations relative to mean scores indicate considerable, but acceptable, variability in student performance on the Math Procedures subtest of the SAT-10, which may be related to student completion of test items. Although students were given 80 minutes to complete the SAT-10, not all students completed every item. An examination of SAT-10 item scores reveals most of the students who did not complete all items completed proportionately fewer Math Procedures items than Math Problem Solving items. Consequently, the variation in Math Procedures performance may be at least partially related to insufficient time to complete the subtest, which is reinforced by positive skew in the distribution of scores. However, examination of distributions indicates normal distributions for the Math Problem Solving subtest and SAT-10 total score. The Algebra Composite distribution is negatively skewed, indicating students, on average, performed better on algebra items than the SAT-10 as a whole.



Graph 1. Graph of mean scores by measurement occasion. CD = Correct Digits (for M-CBM2 only).

Table 4

Descriptive Statistics for SAT-10 and Algebra Outcomes

Measure	*n*	*Min*	*Max*	*M*	*SD*
Math Procedures	198	0	29	8.75	6.32
Math Problem Solving	198	0	45	24.82	9.24
SAT-10	198	0	74	33.57	13.51
Algebra Composite	198	0	8	5.09	2.16

Descriptive Statistics By Course

CBMs. Descriptive statistics for each measurement occasion of the M-CBM2, Basic Skills, Algebra Foundations, and Translations measures for pre-algebra and algebra classrooms are provided in Table 5. Mean scores and standard deviations follow a pattern similar to the pattern described for all classrooms: With the exception of the Basic Skills measure and the first administration of M-CBM2 and Algebra Foundations, mean scores indicate a positive linear trend over time (see Graph 2).

A comparison of means for pre-algebra and algebra students indicates higher mean performance at all data points for students in algebra classrooms, with one exception: The mean score for the second administration of the Translations measure is considerably higher for the five students who moved from pre-algebra classrooms to algebra classrooms during the study. In addition, an examination of mean scores across measures indicates, relative to students in algebra classrooms, students in pre-algebra classrooms performed better on M-CBM2 (i.e., comparing mean measure scores across groups, relative scores for students in pre-algebra classrooms were highest on M-CBM2). It is possible that mean performance for pre-algebra students is highest (in relative terms) on the M-CBM2 measure, because these students may spend more time receiving instruction or practice with the types of skills assessed by the M-CBM2 measure (e.g., mixed computation with rational numbers), when compared to the skills assessed by the other CBMs (e.g., combining like terms, evaluating and solving equations).

Table 5

Descriptive Statistics for Algebra and Mixed Computation CBM by Time and Course

Measure/	Pre-Algebra					Algebra				
Time	*n*	*Min*	*Max*	*M*	*SD*	*n*	*Min*	*Max*	*M*	*SD*
M-CBM2										
Time 1	63	0	42	19.00	8.07	124	4	61	26.15	10.07
Time 2	67	3	40	21.28	7.93	120	3	66	23.75	9.64
Time 3	67	0	49	20.48	10.22	126	6	68	28.46	11.12
Time 4	69	0	47	21.64	9.90	125	0	74	29.49	12.59
Time 5	67	1	55	25.09	11.90	117	4	87	32.34	13.75
Basic Skills										
Time 1	63	0	15	7.46	3.31	124	1	26	12.49	5.81
Time 2	68	1	16	8.57	3.58	119	2	32	13.20	6.12
Time 3	67	0	12	5.07	3.14	126	1	27	10.80	5.96
Time 4	69	0	18	8.91	4.20	126	0	37	14.71	6.72
Time 5	67	2	20	8.79	4.13	117	1	30	13.59	6.37

Measure/ Time	Pre-Algebra					Algebra				
	n	*Min*	*Max*	*M*	*SD*	*n*	*Min*	*Max*	*M*	*SD*
Foundations										
Time 1	63	0	19	10.17	4.92	124	0	32	17.47	5.38
Time 2	68	0	26	9.25	4.32	119	0	33	14.87	5.59
Time 3	67	0	23	9.64	4.72	126	4	31	16.87	5.94
Time 4	69	0	22	10.70	5.24	125	0	35	17.94	6.48
Time 5	67	3	27	12.19	5.37	117	4	40	19.42	8.05
Translations										
Time 2	0	-	-	-	-	120	0	27	8.49	7.40
Time 5	5	1	28	17.20	11.95	115	0	32	11.65	7.21

Note. Foundations = Algebra Foundations.

Outcome measures. Descriptive statistics for the SAT-10 subtests (i.e., Math Procedures and Math Problem Solving), SAT-10 total score, and Algebra Composite for pre-algebra and algebra classrooms are provided in Table 6. Similar to mean scores for the full sample, pre-algebra and algebra mean scores indicate higher math performance for students in algebra classrooms relative to students in pre-algebra classrooms. Variation in Math Procedures scores is greater for pre-algebra students than algebra students. It may be that there is a greater range of skill in pre-algebra classrooms due to parameters defining placement in pre-algebra coursework in 8^{th} grade.

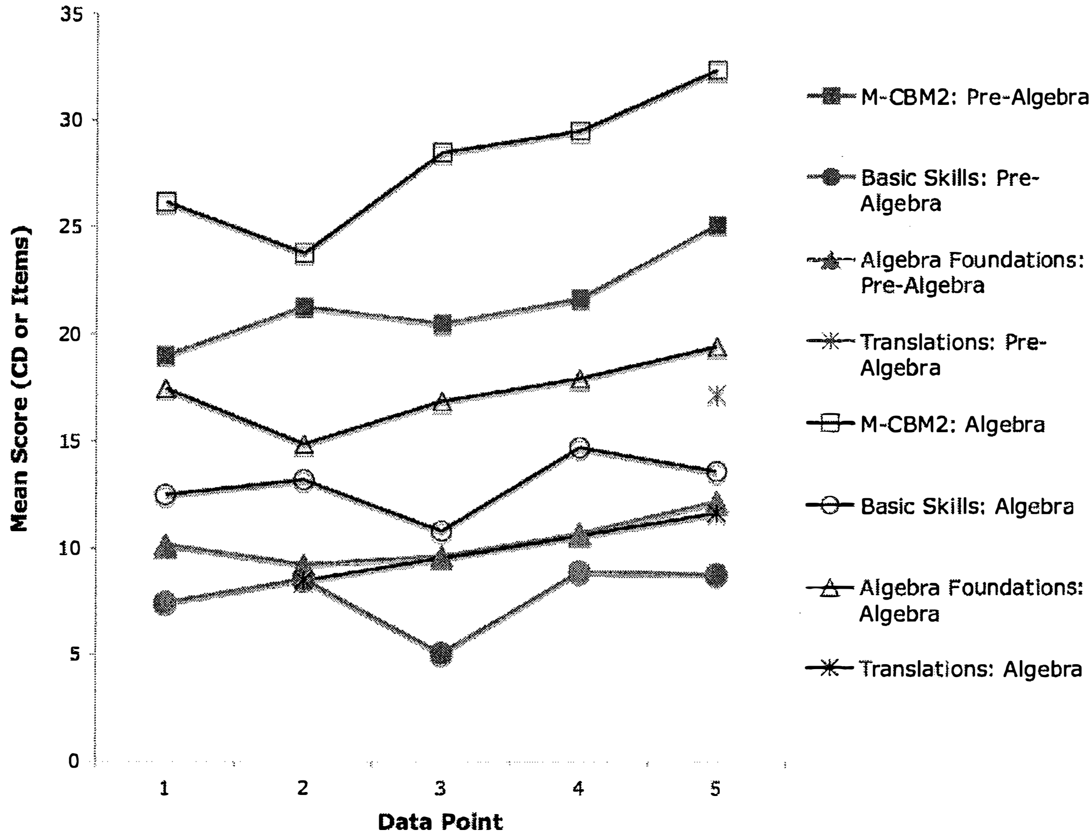
Table 6

Descriptive Statistics for SAT-10 and Algebra Outcomes by Course

	Pre-Algebra					Algebra				
Measure	*n*	*Min*	*Max*	*M*	*SD*	*n*	*Min*	*Max*	*M*	*SD*
Math Procedures	70	0	23	5.39	4.71	128	0	29	10.59	6.34
Math Problem Solving	70	0	37	18.69	7.43	128	0	45	28.17	8.39
SAT-10	70	5	57	24.07	9.90	128	0	74	38.77	12.36
Algebra Composite	70	0	8	4.00	2.12	128	1	8	5.69	1.94

Bivariate Correlations

Reliability of M-CBM2. An examination of correlations between administrations of M-CBM2 indicates moderate relationships between forms of the measure (r = .49 to .74; see Table 7). Correlations indicate relationships are stronger between performances on measurement occasions that occurred closer together, which is expected: Because it is assumed that students will learn over time, performances on measures of the same skills that are administered in closer succession should be more highly correlated than performances on measures administered after a greater lapse of time. However, the relation between proximal administration sessions (e.g., first administration and second administration, second administration and third administration) does not suggest sufficient strength for use in progress monitoring, according to the standards of reliability indicated by Salvia and Ysseldyke (2007).



Graph 2. Graph of mean scores by measurement occasion and course group. CD = Correct Digits (for M-CBM2 only). Translations: Pre-Algebra score at data point 5 is based on $n = 5$.

Validity of M-CBM2. To explore the relationship between student performance on M-CBM2 and outcome measures, bivariate correlations were computed between student performance on the M-CBM2 measure and on both subtests of the SAT-10 (i.e., Math Procedures, Math Problem Solving), as well as total score (see Table 7). In addition, bivariate correlations were computed between student performance on the M-CBM2 measure and student performance on the Algebra Composite, to examine potential differences in the predictive relations between M-CBM2 and outcome measures (see Table 7).

Table 7

Bivariate Correlations Between M-CBM2 and SAT-10 and Algebra Outcomes

	Outcomes				M-CBM2 Data Point				
Data Points	MPRO	MPS	SAT-10	Algebra	1	2	3	4	5
1	.31	.35	.38	.28					
2	.22	.24	.27	.23	.58				
3	.39	.37	.44	.29	.61	.63			
4	.31	.36	.39	.33	.54	.52	.62		
5	.31	.32	.37	.28	.49	.51	.74	.64	

Note. Correlations are based on $n = 187$ (data point 1), $n = 176$ (data point 2), $n = 182$ (data point 3), $n = 183$ (data point 4), and $n = 173$ (data point 5). All correlations are significant at $p < .05$.

Correlation coefficients indicate low to moderate positive relationships ($r = .27$ to .44) between M-CBM2 performance in the fall and SAT-10 performance in the winter. Student performance on M-CBM2 demonstrates low correlations with student performance on the Algebra Composite ($r = .23$ to .33). Because the Algebra Composite contains items that are more purely algebraic in nature, it is not surprising that performance on M-CBM2 demonstrates stronger correlations with performance on SAT-10 total and subtest scores than with performance on the Algebra Composite: As a measure of mixed computation with rational numbers, it is expected that performance on M-CBM2 would be more highly related to performance on another measure more inclusive of general mathematical topics.

According to Salvia & Ysseldyke (2007), acceptable validity coefficients may vary according to types of validity and the purposes for which assessments may be used; however, coefficients above .60 are generally desirable. Using this standard as a means for evaluating validity evidence of M-CBM2 with respect to relations to other variables (AERA, APA, & NCME, 1999), the validity coefficients for M-CBM2 identified in the present study do not appear to have sufficient strength to provide evidence for use of the measure to predict student performance on either the SAT-10 or the Algebra Composite. This finding suggests static fall M-CBM2 scores (i.e., correlations between scores at each data point and outcome measure scores) may not sufficiently predict student general math or algebra performance for students in the middle of 8th grade, which limits the utility of the measure as a progress monitoring tool in middle school.

Reliability of Basic Skills, Algebra Foundations, and Translations. An examination of correlations between administrations of algebra CBM indicates strong relationships between alternate forms (see Tables 8-10). Consistent with the findings for M-CBM2, forms that were administered in closer succession demonstrate greater strength in association. For the Basic Skills and Algebra Foundations measures, correlations between forms are strong (r = .75 to .85 and r = .74 to .81, respectively), which indicates alternate forms are measuring the same behavior. In fact, correlations between successive administrations (e.g., first administration correlated with second administration, second administration correlated with third administration) using single forms indicate high levels of reliability (r = .80 or greater across forms) across nearly all comparisons, which provides strong evidence for the use of the measures to reliably capture the same student behaviors across time. For the Translations measure, alternate forms correlate only moderately (r = .54). However, these forms were administered nearly eight weeks apart, thus we would expect to see a reduced relationship between student performances across the two forms. Consequently, the correlation between the two administrations of Translations cannot serve as an indicator of the reliability of the measure.

Validity of Basic Skills, Algebra Foundations, and Translations. To explore relationships between algebra CBM and general math outcome measures, bivariate correlations were computed for student performance on each algebra CBM (i.e, Basic Skills, Algebra Foundations, and Translations) with both subtests of the SAT-10 (i.e., Math Procedures, Math Problem Solving, and SAT-10 total score). Additional bivariate correlations were computed between algebra CBMs and the Algebra Composite, to explore relations between student performances on algebra CBM and a measure of algebra outcomes (see Tables 8-10).

Table 8

Bivariate Correlations Between Basic Skills and SAT-10 and Algebra Outcomes

	Outcomes				Basic Skills Data Point				
Data Points	MPRO	MPS	SAT-10	Algebra	1	2	3	4	5
1	.42	.64	.63	.54					
2	.41	.61	.60	.50	.84				
3	.48	.60	.63	.50	.80	.85			
4	.38	.61	.60	.46	.78	.81	.81		
5	.37	.56	.56	.46	.72	.75	.78	.80	

Note. Correlations are based on n = 187 (data point 1), n = 176 (data point 2), n = 182 (data point 3), n = 184 (data point 4), and n = 173 (data point 5). All correlations are significant at $p < .01$.

Bivariate correlations between student performances on algebra CBM and general math outcomes indicate moderate to strong predictive relationships for Basic Skills and Algebra Foundations and moderate predictive relationships for Translations. Student performances on Basic Skills and Algebra Foundations measures were more strongly correlated with performances on the Math Problem Solving subtest (r = .56 to .64 and r = .59 to .66, respectively), and SAT-10 total (r = .56 to .63 and r = .58 to .64), respectively), than they were with the Math Procedures subtest (r = .37 to .48 and r = .37 to .46, respectively). Performance on the Translations measure was moderately correlated with performance on the Math Problem Solving subtest (r = .43 and .46) and

SAT-10 total (r = .25 and .38); however the predictive relationship between student performance on the Translations measure and student performance on the Math Procedures subtest (r = -.09 and .13) was negligible and not significant (p > .05).

Table 9

Bivariate Correlations Between Algebra Foundations and SAT-10 and Algebra Outcomes

	Outcomes				Algebra Foundations Data Point				
Data Points	MPRO	MPS	SAT-10	Algebra	1	2	3	4	5
1	.44	.64	.64	.53					
2	.42	.66	.64	.56	.75				
3	.46	.60	.63	.56	.81	.80			
4	.37	.59	.58	.52	.74	.76	.80		
5	.41	.63	.63	.56	.75	.79	.80	.80	

Note. Correlations are based on n = 187 (data point 1), n = 176 (data point 2), n = 182 (data point 3), n = 183 (data point 4), and n = 173 (data point 5). All correlations are significant at p < .01.

The relations between student performance on algebra CBM and the Algebra Composite are moderate and positive. Performance on the Basic Skills measure was related to performance on the Algebra Composite (r = .46 to .54) with greater strength than was indicated in the relations with Math Procedures, but less strength than was indicated in the relations with Math Problem Solving or SAT-10 total scores. Student performance on the Algebra Foundations measure was moderately correlated with student performance on the Algebra Composite (r = .53 to .56), similar to findings for relations with Math Problem Solving and SAT-10 total scores. Correlations between student performance on the Translations measure and the Algebra Composite (r = .44 to .47) and the Translations measure and the Math Problem Solving subtest (r = .43 to .46) indicate stronger relations than were found between student performance on Translations and other general math outcome measures, which may due to the nature of the content assessed by the Translations measure (i.e., conceptual algebra knowledge).

Table 10

Bivariate Correlations Between Translations and SAT-10 and Algebra Outcomes

	Outcomes				Translations Data Point	
Data Points	MPRO	MPS	SAT-10	Algebra	2	5
2	.13	.46*	.38*	.47*		
5	-.09	.43*	.25*	.44*	.54*	

Note. Correlations are based on n = 116, for both data points. *Correlations are significant at $p < .01$.

The reported bivariate correlations indicate stronger relations between variables that can be considered similarly complex: Considering the CBMs examined in this study in a hierarchy of perceived task complexity, M-CBM2 assesses skills taught prior to topics in algebra, Basic Skills is considered the most basic of the algebra measures, Algebra Foundations is representative of typical algebra content, and Translations—the most complex of the measures—assesses advanced conceptual thinking in an initial algebra course. In this regard, it is expected that student performance on the Basic Skills measure would be less related to performance on the Algebra Composite, while student performance on the Algebra Foundations and Translations measures would be more highly related to student performance on the Algebra Composite. Conversely, it is expected that student performance on the Algebra Foundations and Translations measure would be less related to performance on a measure of math calculation (e.g., Math Procedures), while student performance on Basic Skills and M-CBM2 would demonstrate increased relations to general math outcomes. The hypotheses were largely confirmed by the present study.

Using this recommendations of Salvia and Ysseldyke (2007) as a standard for validity evidence based on relations to other variables (AERA, APA, & NCME, 1999), Basic Skills and Algebra Foundations approach providing sufficient evidence of validity for predicting SAT-10 performance. Following the same validity principles, student performance on Algebra Foundations may be valid for predicting an algebra outcome. Still, additional research related to item and content composition of Algebra Foundations and a more established algebra outcome measure may allow for increased understanding of the skills assessed and overall utility of the measure. The Translations measure may also be valid for predicting an algebra outcome; however more research with additional forms would be needed to further examine the findings of this study.

Hierarchical Linear Modeling

As identified in the data analysis section of the last chapter, HLM was used to answer study research questions three, four, and five for the M-CBM2, Basic Skills, and Algebra Foundations measures. For research question three, a 3-level model (i.e., level 1 = time, level 2 = students, level 3 = classrooms) was used to examine growth rates for the measures. For research question four, residual values for slopes and intercepts obtained from the 3-level model were utilized in a second, 2-level model to explore growth rates with respect to SAT-10 and Algebra Composite scores. For research question five, course (i.e., pre-algebra or algebra) was added as a predictor at the classroom level to examine any differential effects on outcome variables for students in pre-algebra or algebra classrooms with respect to research questions three and four. Relevant modeling decisions and results are discussed in the context of remaining research questions in the following sections.

3-Level Growth Model

Three parallel, three-level HLM models were used to account for the time of administration of the measures, individual student variation, and possible classroom level influences with respect to growth rates for the M-CBM2, Basic Skills, and Algebra Foundations measures. The outcome variable for each model was student scores on the measure of interest in the model (i.e., M-CBM2, Basic Skills, or Algebra Foundations), and time was a predictor in each model.

An examination of level two and level three variance components revealed significant variability ($p < .001$) at both levels for each measure model, which provided further support for the use of HLM to analyze the results of this study. Calculation of the unconditional intraclass correlation coefficient (ICC) provided an estimate of the proportion of variance attributable to time (level 1), individual students (level 2) and classrooms (level 3). Specifically, 73% of the variance in measure scores on Algebra Foundations, 74% of the variance in measure scores on Basic Skills, and 55% of the variance in measure scores on M-CBM2 was attributable to individuals and classrooms, while 27%, 26%, and 45% of the variance in measure scores was attributable to time, respectively. The fully unconditional, or null, model (i.e., including no predictors) was:

Level-1 Model

$$Y_{tij} = \pi_{0ij} + e_{tij}$$

Level-2 Model

$$\pi_{0ij} = \beta_{00j} + r_{0ij}$$

Level-3 Model

$$\beta_{00j} = \gamma_{000} + \mu_{00j}$$

where Y_{tij} was the ith student's score on the measure of interest at the tth time point in classroom j, π_{0ij} was the predicted initial status on the measure for student i in classroom j, and e_{tij} was the ith person's residual at time t in classroom j. At level two, β_{00j} represented the predicted mean initial status in classroom j and r_{0ij} was the residual for student i from that mean. At level three, γ_{000} represented the mean initial status across all classrooms and μ_{00j} represented the residual for classroom j.

Modeling decisions. To develop an unconditional growth model for each measure, a time predictor was included in the model. Because the time that lapsed between each testing session was not consistent across all three schools for the first three administration sessions, it was necessary to model differences in administration timing to support accurate analyses and interpretation of results. Several time models were considered for the three-level HLM model. Two of the models captured time as a raw construct corresponding to the data point: one model employed time forward (e.g., data point 1, 2, 3, 4, 5), while the other used time in reverse (e.g., data point 5, 4, 3, 2, 1). The remaining two models specified the time predictor with more precision, corresponding to the number of weeks in the data collection timeline (e.g., week 1, 2, 3, 4, 5, 6, 7, 8) or weeks in reverse (e.g., week 8, 7, 6, 5, 4, 3, 2, 1). Comparing each time model (i.e., time, time in reverse, weeks, or weeks in reverse) to the unconditional, or null model (i.e., the model containing the outcome, but no time or other predictors) for each measure, deviance statistics were examined for significance as an indicator of model fit.

Three factors were of interest with respect to determining the best time predictor model: (a) choosing a time predictor model that would allow the researcher to answer research questions, (b) identifying the time predictor model containing the smallest deviance statistic (i.e., the best model fit), and (c) choosing one time model for the three CBM measures modeled (i.e., M-CBM2, Basic Skills, and Algebra Foundations) to support consistent interpretation of results. With respect to the first criterion, research questions dictated examination of growth rates after the inclusion of initial skill, which ruled out "reverse" models (i.e., because, reverse models would constrain the intercept as final skill). Regarding the second criterion, the "weeks" model provided the best fit for the Basic Skills and Algebra Foundations measures, but the "time" model provided the best fit for the M-CBM2 measure. Closer examination of deviance statistics revealed that the difference between "time" and "weeks" models for M-CBM2 (i.e., time minus weeks = 6791.36 – 6788.74 = 2.62) and Basic Skills (i.e., weeks minus time = 5288.62 – 5284.74 = 3.88) was relatively small, while the difference between "time" and "weeks" models for Algebra Foundations was fairly large (i.e., weeks minus time = 5421.78 – 5399.23 = 22.55). All reported deviance statistics were significant at $p < .001$.

Using the three stated criteria for choosing a consistent model to support interpretation across measures, the "weeks" model was used for analysis. As a product of choosing an appropriate time predictor model, HLM models were created for each of the three measures (i.e., M-CBM2, Basic Skills, and Algebra Foundations), where weeks were included as a predictor of measure scores. Reliability estimates were explored across each of the measure models to examine model functioning. All measure models indicated strong reliability for intercepts (r = .62 - .88), and low to moderate reliability for slopes (r = .10 - .54). In general, all models indicate they were functioning at least as well at the classroom level as they were at the individual level.

Random effects for intercepts and slopes were included in each model at levels 2 and 3 to examine systematic variation in measure scores as a result of included predictors (see Tables 11-13). Interpretation of variance components for the random effects revealed significant ($p < .08$) unexplained variance at levels two and three for intercepts and slopes, for all measures. Pseudo-R^2 statistics were calculated to allow for examination of the proportion of variance in measure scores attributable to the inclusion of the time predictor, relative to the null model. For the M-CBM2 model, the pseudo-R^2 statistic indicated that 22% of the total variance originally attributable to individuals and classrooms was explained by including the time predictor. For the Basic Skills model, the pseudo-R^2 statistic indicated that 8% of the total variance originally attributable to individuals and classrooms was explained by including the time predictor. For the Algebra Foundations model, the pseudo-R^2 statistic indicates that 20% of the total variance originally attributable to individuals and classrooms was explained by including the time predictor.

Conditional ICCs were also calculated, to examine the proportion of variance explained at each level for each measure model after including the time predictor (see Tables 11-13). For the M-CBM2 model, 48% of the variance in measure scores was attributable to time, 43% of the variance was attributable to individual students, and 9% of the variance was attributable to classrooms. For the Basic Skills model, 26% of the variance in measure scores was attributable to time, 55% of the variance was attributable to individual students, and 19% of the variance was attributable to classrooms. For the

Algebra Foundations model, 28% of the variance in measure scores was attributable to time, 41% of the variance was attributable to individual students, and 31% of the variance was attributable to classrooms. These ICCs indicate the amount of variance in measure scores that are attributable to time, individual students, and classrooms, when data is examined using the unconditional growth model. Because no predictors were of interest to at the student or classroom levels to answer initial research questions, the final HLM growth model was:

Level-1 Model

$$Y_{tij} = \pi_{0ij} + \pi_{1ij}(\text{WEEKS})_{tij} + e_{tij}$$

Level-2 Model

$$\pi_{0ij} = \beta_{00j} + r_{0ij}$$

$$\pi_{1ij} = \beta_{10j} + r_{1ij}$$

Level-3 Model

$$\beta_{00j} = \gamma_{000} + \mu_{00j}$$

$$\beta_{10j} = \gamma_{100} + \mu_{10j}$$

where Y_{tij} was the ith student's score on the measure of interest at the tth time point in classroom j, π_{0ij} was the predicted initial status on the measure for student i in classroom j, π_{1ij} was the rate of change in measure scores across weeks, and e_{tij} was the ith person's residual at time t in classroom j. At level two, β_{00j} represented the predicted mean initial status in classroom j, r_{0ij} was the residual for student i from that mean, β_{10j} was the rate of change in measure scores across weeks associated with classroom j, and r_{1ij} was the residual for student i from that rate of change. At level three, γ_{000} represented the mean initial status across all classrooms, μ_{00j} represented the residual for classroom j, γ_{100} was the mean rate of change for all classrooms, and μ_{10j} was the residual for classroom j from the mean rate of change.

Table 11

Fixed Effects and Variance Components for M-CBM2, Final Growth Model

Fixed Effect	Coefficient	*se*	*t* Ratio	*p*
Predicted initial status	22.43	1.12	20.04	.000
Growth rate	0.94	0.17	5.69	.000

Random Effect	Variance Component	*df*	X^2	*p*
Time (level 1)	47.01			
Students (level 2)	41.33	188	493.03	.000
Weeks	0.97	188	284.36	.000
Classrooms (level 3)	8.40	9	39.23	.000
Weeks	0.59	9	23.20	.006

Variance Decomposition (Percentage by level)	
Level 1	47.8
Level 2	43.0
Level 3	9.2

How much growth can be expected? Research question three requires an examination of slopes and p-values for the unconditional growth model for each measure (i.e., M-CBM2, Basic Skills, and Algebra Foundations). Fixed effects indicate growth in measure scores is significant at $p < .05$ (see Tables 11-13). For M-CBM2, it is predicted that the average 8^{th} grade student will score 22.43 correct digits on an M-CBM2 probe given at the beginning of fall; for each week that passes, the average 8^{th} grade student's performance is expected to increase 0.94 digits. For Basic Skills, it is predicted that the average 8^{th} grade student will answer correctly 9.73 items on a Basic Skills probe given at the beginning of fall; for each week that passes, the average 8^{th} grade student's performance is expected to grow by 0.22 items. For Algebra Foundations, it is predicted that the average 8^{th} grade student will answer 12.95 items correct on an Algebra Foundations probe given at the beginning of fall; for each week that passes, the average 8^{th} grade student's performance is expected to increase .39 items. Multiplying the growth rate by the number of weeks in the study allows for a determination of total growth in student performance across the fall for each measure. Data was collected in this study for eight weeks. On average, during this study student performance grew 7.52 correct digits on M-CBM2, 1.76 items correct on Basic Skills, and 3.12 items correct on Algebra Foundations. These growth rates may be used to provide educators with an indication of expected growth on the studied measures in the fall of 8^{th} grade.

Table 12

Fixed Effects and Variance Components for Basic Skills, Final Growth Model

Fixed Effect	Coefficient	*se*	*t* Ratio	*p*
Predicted initial status	9.73	0.90	10.80	.000
Growth rate	0.22	0.07	3.14	.013

Random Effect	Variance Component	*df*	X^2	*p*
Time (level 1)	8.84			
Students (level 2)	18.66	188	942.20	.000
Weeks	0.04	188	217.75	.068
Classrooms (level 3)	6.52	9	66.69	.000
Weeks	0.03	9	23.38	.006

Variance Decomposition (Percentage by level)	
Level 1	25.9
Level 2	54.9
Level 3	19.2

Table 13

Fixed Effects and Variance Components for Algebra Foundations, Final Growth Model

Fixed Effect	Coefficient	*se*	*t* Ratio	*p*
Predicted initial status	12.95	1.10	11.72	.000
Growth rate	0.39	0.07	6.05	.000

Random Effect	Variance Component	*df*	X^2	*p*
Time (level 1)	10.03			
Students (level 2)	14.26	188	695.76	.000
Weeks	0.18	188	274.45	.000

Random Effect	Variance Component	*df*	X^2	*p*
Classrooms (level 3)	10.77	9	129.09	.000
Weeks	0.01	9	15.79	.071

Variance Decomposition (Percentage by level)	
Level 1	28.4
Level 2	41.0
Level 3	30.6

2-Level Empirical Bayes Coefficient Model

To explore whether growth rates predicted general math or algebra outcomes above and beyond initial skills, two-level models were used to examine growth rates and predicted initial status for measures included in the study with respect to SAT-10 and Algebra Composite scores. Two-level unconditional, or null, models were created for SAT-10 and Algebra Composite outcomes. In the 2-level Empirical Bayes Coefficient (EC) models, level one represented individual students and level two represented classrooms. Reliability estimates for each model were explored. Both unconditional measure models revealed strong reliability (SAT-10, $r = .85$; Algebra Composite, $r = .81$). An examination of the variance components for both null models indicated significant variability ($p < .001$) in student outcomes as a function of classrooms (level 2), which provided further support for the use of HLM to analyze the results of this study.

Calculation of the unconditional intraclass correlation coefficients (ICC) allowed for examination of the proportion of variance attributable to individual students (level 1) and classrooms (level 2). Specifically, 73% of the variance in SAT-10 scores was attributable to individual students, while 27% was attributable to classrooms. 77% of the variance in Algebra Composite scores was associated with individual students, while 23% of the variance was associated with classrooms. The unconditional, or null, model was

Level-1 Model

$$Y_{ij} = \beta_{0j} + r_{ij}$$

Level-2 Model

$$\beta_{0j} = \gamma_{00} + \mu_{0j}$$

where Y_{ij} was the ith student's score on the outcome of interest (i.e., SAT-10 or Algebra Composite) in classroom j, β_{0j} represented the average score on the outcome measure, and r_{ij} was the residual for student i from that mean. At level two, γ_{00} represented the mean score on the outcome of interest across all classrooms and μ_{0j} represented the residual for classroom j.

Modeling decisions. Empirical Bayes Coefficients (ECs) for each final three-level measure model were extracted from 3-level model analyses for use in the two-level EC models. Using these values, an EC model was created for each CBM with more than two measurement occasions (i.e., M-CBM2, Basic Skills, and Algebra Foundations), for each of the outcomes of interest (i.e., SAT-10 and Algebra Composite). In all, six EC models are included in this analysis.

At level one, predictors added to EC models consisted of Empirical Bayes Coefficients (i.e., residual values plus fitted values) for slopes and intercepts for each student with respect to each measure. The EC values were added at level one as grand-mean centered predictors to constrain the intercept to be the mean outcome measure score for students with average predicted CBM initial status and average CBM slope. The EC values for intercepts and slopes were selected for use in the model because (a) the coefficients provide direct information about individual students' predicted initial status and slopes on each CBM, (b) EC values control for imbalances in sample size, (c) EC values respond well to the presence of missing data, and (d) EC values provide a conservative estimate of variance. No additional predictors were included at levels one or two in the initial EC models. The final (conditional, fixed effects) EC model was

Level-1 Model

$$Y_{ij} = \beta_{0j} + \beta_{1j}(\textbf{\textit{EC INTERCEPT}}) + \beta_{2j}(\textbf{\textit{EC SLOPE}}) + r_{ij}$$

Level-2 Model

$$\beta_{0j} = \gamma_{00} + \mu_{0j}$$

$$\beta_{1j} = \gamma_{10}$$

$$\beta_{2j} = \gamma_{20}$$

where Y_{ij} was the ith student's score on the outcome of interest (i.e., SAT-10 or Algebra Composite) in classroom j, β_{0j} represented the average score on the outcome measure for students with average slope and intercept EC values for the CBM of interest, β_{1j} represented the change in the outcome measure corresponding with a change in intercept EC value for a student in classroom j, β_{2j} was the change in the outcome measure corresponding with a change in the slope EC value for a student in classroom j, and r_{ij} was the residual for student i in classroom j. At level two, γ_{00} represented the mean outcome score across all classrooms, γ_{10} was the mean intercept EC value for the CBM of interest, γ_{20} was the mean slope EC value for the CBM of interest, and μ_{0j} represented the residual for classroom j.

Random effects were explored at level two, in part because significant variance was identified at the classroom level in the conditional, fixed effects model; however, simultaneously including random effects for the slope and the intercept yielded nonsignificant random variance and low degrees of freedom, suggesting the number of units at level two may have been insufficient to estimate random variance for the parameters. In addition, including random effects did not affect the significance of the fixed effects, which will be a primary source of data to answer research questions four and five. Consequently, variance was constrained as fixed for each measure model. Because random variance was not included in any model, the fixed effects model became the final model, thus pseudo-R^2 statistics could not be computed.

Reliability estimates were explored across each of the measure models to examine model functioning. Models containing M-CBM2 or Basic Skills demonstrated strong reliability (r = .66 to .79) for both SAT-10 and Algebra Composite outcomes. The model containing Algebra Foundations and SAT-10 outcomes demonstrated poor reliability (r = .01), while the model containing Algebra Foundations and Algebra Composite scores demonstrated moderate reliability (r = .42). This result suggests, with the exception of Algebra Foundations with SAT-10 outcomes, model intercept reliability is acceptable.

Conditional ICCs were calculated to determine the amount of variance in outcome measures associated with individual students and classrooms, after including EC values for predicted initial skills and slopes in each model (see Tables 14-19). In the model containing SAT-10 outcomes and M-CBM2 coefficients, conditional ICCs indicated 79% of the variance in SAT-10 outcomes was associated with individual students, while 21% of the variance was associated with classrooms. For the model containing SAT-10 outcomes and Basic Skills coefficients, 89% of the variance in SAT-10 outcomes was associated with individual students, while 11% was associated with classrooms. For the model containing SAT-10 outcomes and Algebra Foundations coefficients, > 99% of the variance in SAT-10 outcomes was associated with individuals, while < 1% was associated with classrooms. For Algebra Composite scores, the inclusion of M-CBM2 coefficients in the model indicated 83% of the variance in Algebra Composite scores was associated with individual students, while 17% was associated with classrooms. For the model containing Algebra Composite scores and Basic Skills coefficients, 91% of the variance was associated with individual students, while 9% was attributable to classrooms. For the model containing Algebra Composite scores and Algebra Foundations coefficients, 96% of the variance in Algebra Composite scores was associated with individual students, while 4% was associated with classrooms. Taken together these conditional ICCs indicate that more than three fourths of the variance in general math and algebra outcome measures is associated with individual student differences, while less than one fourth of the variance in general math and algebra outcome measures is associated with classrooms.

Comparing models across outcomes, the Algebra Composite, Algebra Foundations model demonstrated more variance at the classroom level when compared to the SAT-10, Algebra Foundations model, while for other measure models, variance trended in the opposite direction. Corresponding to student performance on the Algebra Foundations measure, classrooms explain more variance in algebra outcomes than they do in general math outcomes. It was hypothesized this difference could be due to the presumed instructional focus of the classroom (i.e., pre-algebra or algebra course).

Table 14

Fixed Effects and Variance Components for M-CBM2, Empirical Bayes Coefficient Model with SAT-10 Outcomes

Fixed Effect	Coefficient	*se*	*t* Ratio	*p*
SAT-10[a]	32.65	1.97	16.57	.000
Intercept residual	0.68	0.22	3.15	.002
Slope residual	0.48	1.62	0.30	.766

Random Effect	Variance Component	*df*	X^2	*p*
Students (level 1)	117.87			
Classrooms (level 2)	30.67	9	56.47	.000

Variance Decomposition (Percentage by level)	
Level 1	79.4
Level 2	20.6

[a] SAT-10 represents the average SAT-10 score for a student with an average intercept and average slope EC value.

Table 15

Fixed Effects and Variance Components for M-CBM2, Empirical Bayes Coefficient Model with Algebra Composite Outcomes

Fixed Effect	Coefficient	*se*	*t* Ratio	*p*
Algebra Composite[a]	4.95	0.30	16.50	.000
Intercept residual	0.07	0.04	1.82	.071
Slope residual	0.21	0.30	0.70	.484

Random Effect	Variance Component	*df*	X^2	*p*
Students (level 1)	3.33			
Classrooms (level 2)	0.68	9	48.43	.000

Variance Decomposition (Percentage by level)	
Level 1	83.0
Level 2	17.0

[a] Algebra Composite represents the average Algebra Composite score for a student with an average intercept and average slope EC value.

Does slope predict outcomes, above and beyond initial skills? Research question four requires an examination of fixed effect coefficients and their corresponding *p*-values. The EC values associated with intercepts contribute significantly to SAT-10 and Algebra Composite scores (see Tables 14-19), where a 1-unit increase in predicted initial skill is associated with increased SAT-10 and Algebra Composite scores. EC values for slopes indicate small increases in SAT-10 and Algebra Composite scores when M-CBM2 slopes increase; moderate increases in SAT-10 and small decreases in Algebra Composite scores when Basic Skills slopes increase; and moderate decreases in SAT-10 and small increases in Algebra Composite scores when Algebra Foundations slopes increase. However, fixed effect coefficients for slope were not significant and cannot be accurately interpreted.

These results indicates that slope does not predict SAT-10 and Algebra Composite scores above and beyond initial skills for any CBM included in this study. Overall, results indicate students who have higher performance on M-CBM2, Basic Skills, or Algebra Foundations probes at the beginning of fall of 8th grade (i.e., higher initial skills) will score higher on general math or algebra outcome measures. Measure slopes should be examined over a longer period of time to gain more insight about their utility for predicting general math and algebra outcomes in middle school mathematics, above and beyond initial skills.

Table 16

Fixed Effects and Variance Components for Basic Skills, Empirical Bayes Coefficient Model with SAT-10 Outcomes

Fixed Effect	Coefficient	*se*	*t* Ratio	*p*
SAT-10[a]	33.37	1.32	25.20	.000
Intercept residual	1.58	0.26	6.13	.000
Slope residual	2.28	8.76	0.26	.795

Random Effect	Variance Component	*df*	X^2	*p*
Students (level 1)	93.37			
Classrooms (level 2)	11.59	9	28.31	.001

Variance Decomposition (Percentage by level)	
Level 1	89.0
Level 2	11.0

[a] SAT-10 represents the average SAT-10 score for a student with an average intercept and average slope EC value.

Table 17

Fixed Effects and Variance Components for Basic Skills, Empirical Bayes Coefficient Model with Algebra Composite Outcomes

Fixed Effect	Coefficient	*se*	*t* Ratio	*p*
Algebra Composite[a]	5.05	0.22	23.35	.000
Intercept residual	0.22	0.04	5.06	.000
Slope residual	-0.52	1.47	-0.36	.721

Random Effect	Variance Component	*df*	X^2	*p*
Students (level 1)	2.93			
Classrooms (level 2)	0.29	9	25.19	.003

Variance Decomposition (Percentage by level)	
Level 1	91.0
Level 2	9.0

[a] Algebra Composite represents the average Algebra Composite score for a student with an average intercept and average slope EC value.

Table 18

Fixed Effects and Variance Components for Algebra Foundations, Empirical Bayes Coefficient Model with SAT-10 Outcomes

Fixed Effect	Coefficient	*se*	*t* Ratio	*p*
SAT-10[a]	33.57	0.68	49.45	.000
Intercept residual	2.25	0.27	8.25	.000
Slope residual	-4.63	3.78	-1.23	.222

Random Effect	Variance Component	*df*	X^2	*p*
Students (level 1)	90.53			
Classrooms (level 2)	0.03	9	4.95	>.500

Variance Decomposition (Percentage by level)	
Level 1	100.0
Level 2	0.0

[a] SAT-10 represents the average SAT-10 score for a student with an average intercept and average slope EC value.

Table 19

Fixed Effects and Variance Components for Algebra Foundations, Empirical Bayes Coefficient Model with Algebra Composite Outcomes

Fixed Effect	Coefficient	*se*	*t* Ratio	*p*
Algebra Composite[a]	5.08	0.16	31.63	.000
Intercept residual	0.27	0.06	4.77	.000
Slope residual	0.08	0.71	0.11	.915

Random Effect	Variance Component	*df*	X^2	*p*
Students (level 1)	2.62			
Classrooms (level 2)	0.11	9	14.40	.108

Variance Decomposition (Percentage by level)	
Level 1	95.9
Level 2	4.1

[a] Algebra Composite represents the average Algebra Composite score for a student with an average intercept and average slope residual

Differences for Algebra and Pre-Algebra Classrooms

The 3-level final growth models and the 2-level EC models were used to answer research question five. To determine whether there were differences in measure growth rates for pre-algebra or algebra students during the fall, course (i.e., algebra or pre-algebra) was added as a predictor to the 3-level unconditional growth models. To explore whether there were differences in the predictive relation between growth rates and algebra or general math outcomes for pre-algebra and algebra students after taking into account predicted initial skills, course was also added as a level two predictor to each 2-level EC model.

3-Level growth model. The final (unconditional) growth model containing time predictor and measure scores at each measurement occasion indicated significant variability at levels two and three. Course was examined as an additional predictor at level three in an attempt to explain classroom variance in measure scores and explore differences for these groups with reference to research questions. Adding course as a predictor at level three, the content growth model was

Level-1 Model

$$Y_{tij} = \pi_{0ij} + \pi_{1ij}(\text{WEEKS})_{tij} + e_{tij}$$

Level-2 Model

$$\pi_{0ij} = \beta_{00j} + r_{0ij}$$

$$\pi_{1ij} = \beta_{10j} + r_{1ij}$$

Level-3 Model

$$\beta_{00j} = \gamma_{000} + \gamma_{001}(\text{COURSE}) + \mu_{00j}$$

$$\beta_{10j} = \gamma_{100} + \gamma_{101}(\text{COURSE}) + \mu_{10j}$$

where Y_{tij} was the ith student's score on the CBM of interest at the tth time point in classroom j, π_{0ij} was the predicted initial status on the measure for student i in classroom

j, π_{1ij} was the rate of change in measure scores across weeks, and e_{tij} was the ith person's residual at time t in classroom j. At level two, β_{00j} represented the predicted mean initial status in classroom j, r_{0ij} was the residual for student i from that mean, β_{10j} was the rate of change in measure scores across weeks associated with classroom j, and r_{1ij} was the residual for student i from that rate of change. At level three, γ_{000} represented the mean initial status across all pre-algebra classrooms, γ_{001} indicated the difference in the intercept associated with enrollment in an algebra course, μ_{00j} represented the residual for classroom j, γ_{100} was the mean rate of change for all pre-algebra classrooms, γ_{101} indicated the difference in the slope associated with algebra classrooms, and μ_{10j} was the residual for classroom j from the mean rate of change.

Pseudo-R^2 statistics were calculated to allow for examination of the proportion of variance in CBM scores explained by including course as a predictor, relative to the unconditional growth model. For all models, the pseudo-R^2 statistic indicated that little to no additional variance in growth rates was explained by adding course as a predictor at level three. However, an examination of ICCs indicates that course explained the bulk of the variance in growth rates attributable to classrooms. When course was included as a predictor at level three, variance associated with classrooms decreased by approximately 6% for M-CBM2, variance associated with classrooms decreased by nearly 17% for Basic Skills, and variance associated with classrooms decreased by over 20% for Algebra Foundations (see Tables 20-22). It's possible that these decreases in variance are associated with the type of skills the measures assess. For example, because Algebra

Foundations assesses more complex algebra-focused skills than the other measures, including course as a predictor at the classroom level may have explained the bulk of the variance associated with classrooms for this measure.

Table 20

Fixed Effects and Variance Components for M-CBM2, Course Growth Model

Fixed Effect	Coefficient	*se*	*t* Ratio	*p*
Predicted initial status	14.05	2.66	5.30	.000
Course[a]	5.34	1.57	3.39	.011
Growth rate	0.24	0.50	0.48	.642
Course[a]	0.45	0.30	1.52	.166

Random Effect	Variance Component	*df*	X^2	*p*
Time (level 1)	47.00			
Students (level 2)	41.06	188	493.13	.000
Weeks	0.98	188	284.42	.000
Classrooms (level 3)	2.21	8	18.73	.016
Weeks	0.06	8	19.47	.013

Variance Decomposition (Percentage by level)	
Level 1	51.5
Level 2	46.0
Level 3	2.5

[a] Course: 0 = Pre-algebra classrooms, 1 = Algebra classrooms

The course growth model indicates that, when course is added as a predictor in the model, intercepts predict CBM scores for algebra students ($p < .05$), but intercepts for pre-algebra students (with the exception of M-CBM2) and slopes for pre-algebra and algebra students are not significantly different from zero (see Tables 20-22). Coefficients in the course growth model indicate, on average, students in algebra classrooms have a higher predicted initial status than pre-algebra students in the fall on all three CBMs.

Specifically, students in algebra classrooms are expected to score 5.34 correct digits higher than students in pre-algebra classrooms on M-CBM2 probes at the beginning of the fall. Students in algebra classrooms are also expected to score 4.98 items higher for Basic Skills and 6.52 items higher for Algebra Foundations, when compared to pre-algebra students at the beginning of fall. For pre-algebra students taking M-CBM2 probes, predicted initial status was also a significant predictor of measure scores: Pre-algebra students are expected to begin the fall scoring 14.05 correct digits, and algebra students are expected to score 19.39 (i.e., 14.05 + 5.34) correct digits at the beginning of fall. After adding course as a predictor in the model, slopes for CBMs decreased overall and were no longer significant. It's possible that slopes are not significant because course confounds slopes for each CBM, because the data collection timeline employed in this study was not long enough to allow students to demonstrate significant growth in pre-algebra or algebra classrooms, or because students in the sample did not improve math skills assessed by the measures studied (e.g., instruction did not lead to measure-relevant learning during the fall). Future studies should incorporate methodology to explore these issues.

Table 21

Fixed Effects and Variance Components for Basic Skills, Course Growth Model

Fixed Effect	Coefficient	*se*	*t* Ratio	*p*
Predicted initial status	2.16	1.57	1.38	.204
Course[a]	4.98	0.93	5.35	.000
Growth rate	0.08	0.23	0.35	.739
Course[a]	0.10	0.14	0.68	.513

Random Effect	Variance Component	*df*	X^2	*p*
Time (level 1)	8.85			
Students (level 2)	18.40	188	941.81	.000
Weeks	0.04	188	217.66	.068
Classrooms (level 3)	0.78	8	16.22	.039
Weeks	0.03	8	21.73	.006

Variance Decomposition (Percentage by level)	
Level 1	31.5
Level 2	65.6
Level 3	2.9

[a] Course: 0 = Pre-algebra classrooms, 1 = Algebra classrooms

Table 22

Fixed Effects and Variance Components for Algebra Foundations, Course Growth Model

Fixed Effect	Coefficient	*se*	*t* Ratio	*p*
Predicted initial status	2.89	1.64	1.77	.115
Course[a]	6.52	0.98	6.63	.000
Growth rate	0.18	0.22	0.84	.427
Course[a]	0.14	0.13	1.07	.318

Random Effect	Variance Component	*df*	X^2	*p*
Time (level 1)	10.02			
Students (level 2)	14.14	188	695.92	.000
Weeks	0.17	188	274.52	.000
Classrooms (level 3)	1.21	8	23.54	.003
Weeks	0.01	8	13.72	.089

Variance Decomposition (Percentage by level)	
Level 1	39.2
Level 2	56.0
Level 3	4.8

[a] Course: 0 = Pre-algebra classrooms, 1 = Algebra classrooms

2-Level EC Model. Significant variability ($p < .05$) at the classroom level was identified for each of the final EC models containing intercept and slope residuals for M-CBM2, Basic Skills, and Algebra Foundations measures with respect to SAT-10 and

Algebra Composite outcomes, with the exception of the Algebra Foundations, SAT-10 model. Course was examined as an additional predictor at level two in an attempt to explain classroom variance in measure scores and explore possible differences for these groups with reference to research questions. Adding course as a predictor at level two, the course EC model was

Level-1 Model

$$Y_{ij} = \beta_{0j} + \beta_{1j}(\textbf{\textit{EC INTERCEPT}}) + \beta_{2j}(\textbf{\textit{EC SLOPE}}) + r_{ij}$$

Level-2 Model

$$\beta_{0j} = \gamma_{00} + \mu_{0j}$$

$$\beta_{1j} = \gamma_{10} + \gamma_{11}(\text{COURSE})$$

$$\beta_{2j} = \gamma_{20} + \gamma_{21}(\text{COURSE})$$

where Y_{ij} was the ith student's score on the outcome of interest (i.e., SAT-10 or Algebra Composite) in classroom j, β_{0j} represented the average score on the outcome measure for students with average slope and intercept EC values, β_{1j} represented the change in the outcome measure corresponding with a change in intercept EC value for a student in classroom j, β_{2j} was the change in the outcome measure corresponding with a change in the slope EC value for a student in classroom j, and r_{ij} was the residual for student i in classroom j. At level two, γ_{00} represented the mean outcome score across all pre-algebra classrooms, γ_{10} was the mean intercept EC value in pre-algebra classrooms, γ_{11} was the difference in the intercept EC value associated with algebra classrooms, γ_{20} was the mean slope EC value in pre-algebra classrooms, γ_{21} was the difference in the slope EC value associated with algebra classrooms, and μ_{0j} represented the residual for classroom j.

Pseudo-R^2 statistics were calculated to allow for examination of the proportion of variance in SAT-10 and Algebra Composite outcomes explained by including course as a predictor, relative to the final (conditional) EC model. For the M-CBM2 and Algebra Foundations models, the pseudo-R^2 statistic indicated that roughly zero additional variance in SAT-10 and Algebra Composite outcomes was explained by adding course as a predictor at level two. For the Basic Skills models, the pseudo-R^2 statistic indicated that adding course to each model explained an additional 2.4% of the variance in SAT-10 outcomes and 3.1% of the variance in Algebra Composite scores (see Tables 23-28).

By including course as a predictor at level two, it is possible to examine whether EC intercepts and slopes for each measure predict algebra or general math outcomes for students in algebra classrooms compared to students in pre-algebra classrooms. For all measures, slopes did not predict SAT-10 or Algebra Composite scores above and beyond initial skills. However, initial skills significantly predicted SAT-10 and Algebra Composite scores for some measures (see Tables 23-28).

On average, for students in pre-algebra classrooms, predicted initial skills on M-CBM2 and the Algebra Foundations measure contributed significantly to SAT-10 and Algebra Composite scores: for every one-unit increase in the EC intercept value (i.e., predicted initial skills on the CBM), SAT-10 performance increased by 0.98 and 2.48 points, respectively. For every one-unit increase in the EC intercept value, Algebra

Composite performance increased by 0.13 and 0.23 points, respectively ($p < .07$). No significant differences were observed with respect to either algebra or general math outcomes by predicted initial skill on either of the measures for students in algebra classrooms. For the Basic Skills measure, predicted initial skill demonstrated a significant ($p < .07$) predictive relationship with SAT-10 and Algebra Composite scores: In pre-algebra classrooms, for every one unit increase in the in the EC intercept value, SAT-10 and Algebra Composite scores increased by 2.72 and .55 points, respectively. In algebra classrooms, for every one unit increase in the EC intercept value, SAT-10 and Algebra Composite scores increased by 1.35 (i.e., 2.72 - 1.37) and .17 (i.e., 0.55 - 0.38) points, respectively.

These results suggest, on average, initial status on the Basic Skills differentially predicts general math and algebra outcomes for students in pre-algebra and algebra classrooms, where higher predicted initial status in pre-algebra classrooms is associated with increased performance on outcome measures and higher predicted initial status in algebra classrooms results in relatively smaller increases in performance on outcome measures. As noted previously, if the measures included in this study are considered in a hierarchy of difficulty, M-CBM2 assesses the most basic math skills, Basic Skills can be considered a basic measure of algebra skill, and Algebra Foundations assesses core algebra skill. Based on the finding that increases in initial status on the Basic Skills measure contribute to reduced increases in performance on the SAT-10 and Algebra Composite in algebra classrooms relative to pre-algebra classrooms, it's possible that the

Basic Skills measure is more sensitive to differences in skill among students in pre-algebra classrooms when compared to students in algebra classrooms.

Table 23

Fixed Effects and Variance Components for M-CBM2, EC Course Model with SAT-10 Outcomes

Fixed Effect	Coefficient	*se*	*t* Ratio	*p*
SAT-10[a]	33.41	1.89	17.70	.000
Intercept residual	0.98	0.42	2.37	.019
Course[b]	-0.41	0.50	-0.83	.411
Slope residual	1.56	3.18	0.49	.624
Course[b]	-1.61	3.86	-0.42	.676

Random Effect	Variance Component	*df*	X^2	*p*
Students (level 1)	117.42			
Classrooms (level 2)	26.03	9	48.29	.000

Variance Decomposition (Percentage by level)	
Level 1	81.9
Level 2	18.1

[a] SAT-10 represents the average SAT-10 score for a student with an average intercept and average slope residual

[b] Course: 0 = Pre-algebra classrooms, 1 = Algebra classrooms

Table 24

Fixed Effects and Variance Components for M-CBM2, EC Course Model with Algebra Composite Outcomes

Fixed Effect	Coefficient	*se*	*t* Ratio	*p*
Algebra Composite[a]	5.08	0.30	16.96	.000
Intercept residual	0.13	0.07	1.88	.061
Course[b]	-0.08	0.08	-0.99	.322
Slope residual	0.34	0.53	0.64	.523
Course[b]	0.19	0.64	-0.30	.763

Random Effect	Variance Component	*df*	X^2	*p*
Students (level 1)	3.29			
Classrooms (level 2)	0.64	9	44.97	.000

Variance Decomposition (Percentage by level)	
Level 1	83.7
Level 2	16.3

[a] Algebra Composite represents the average Algebra Composite score for a student with an average intercept and average slope residual

[b] Course: 0 = Pre-algebra classrooms, 1 = Algebra classrooms

Table 25

Fixed Effects and Variance Components for Basic Skills, EC Course Model with SAT-10 Outcomes

Fixed Effect	Coefficient	*se*	*t* Ratio	*p*
SAT-10[a]	35.21	1.14	30.91	.000
Intercept residual	2.72	0.68	4.01	.000
Course[b]	-1.37	0.73	-1.89	.060
Slope residual	4.38	23.27	0.19	.851
Course[b]	-2.40	24.52	-0.10	.923

Random Effect	Variance Component	*df*	X^2	*p*
Students (level 1)	91.13			
Classrooms (level 2)	4.98	9	15.12	.087

Variance Decomposition (Percentage by level)	
Level 1	94.8
Level 2	5.2

[a] SAT-10 represents the average SAT-10 score for a student with an average intercept and average slope residual

[b] Course: 0 = Pre-algebra classrooms, 1 = Algebra classrooms

Table 26

Fixed Effects and Variance Components for Basic Skills, EC Course Model with Algebra Composite Outcomes

Fixed Effect	Coefficient	*se*	*t* Ratio	*p*
Algebra Composite[a]	5.29	0.24	22.20	.000
Intercept residual	0.55	0.12	4.40	.000
Course[b]	-0.38	0.13	-2.83	.006
Slope residual	-7.32	4.32	-1.70	.091
Course[b]	7.42	4.60	1.62	.107

Random Effect	Variance Component	*df*	X^2	*p*
Students (level 1)	2.81			
Classrooms (level 2)	0.29	9	24.89	.003

Variance Decomposition (Percentage by level)	
Level 1	90.6
Level 2	9.4

[a] Algebra Composite represents the average Algebra Composite score for a student with an average intercept and average slope residual

[b] Course: 0 = Pre-algebra classrooms, 1 = Algebra classrooms

Table 27

Fixed Effects and Variance Components for Algebra Foundations, EC Course Model with SAT-10 Outcomes

Fixed Effect	Coefficient	*se*	*t* Ratio	*p*
SAT-10[a]	34.44	1.03	33.29	.000
Intercept residual	2.48	0.48	5.15	.000
Course[b]	-0.59	0.80	-0.74	.459
Slope residual	-2.51	6.92	-0.36	.717
Course[b]	-0.08	9.63	-0.01	.994

Random Effect	Variance Component	*df*	X^2	*p*
Students (level 1)	90.33			
Classrooms (level 2)	0.02	9	3.47	>.500

Variance Decomposition (Percentage by level)	
Level 1	100.0
Level 2	0.0

[a] SAT-10 represents the average SAT-10 score for a student with an average intercept and average slope residual

[b] Course: 0 = Pre-algebra classrooms, 1 = Algebra classrooms

Table 28

Fixed Effects and Variance Components for Algebra Foundations, EC Course Model with Algebra Composite Outcomes

Fixed Effect	Coefficient	*se*	*t* Ratio	*p*
Algebra Composite[a]	5.09	0.23	22.52	.000
Intercept residual	0.23	0.10	2.36	.019
Course[b]	0.06	0.15	0.37	.711
Slope residual	1.38	1.26	1.10	.275
Course[b]	-.79	1.75	-1.03	.307

Random Effect	Variance Component	*df*	X^2	*p*
Students (level 1)	2.61			
Classrooms (level 2)	0.12	9	14.94	.092

Variance Decomposition (Percentage by level)	
Level 1	95.6
Level 2	4.4

[a] Algebra Composite represents the average Algebra Composite score for a student with an average intercept and average slope residual

[b] Course: 0 = Pre-algebra classrooms, 1 = Algebra classrooms

Multiple Regression and Paired Samples *t*-test

Because the Translations measure was only administered at two time points during the fall, there is insufficient data to model student growth from student scores on this measure. However, gain, or change, scores can be examined in light of the remaining applicable research questions (i.e., research questions three and four) as an approximation for growth. A paired samples *t*-test was used to identify the amount of gain students demonstrated between early and late fall administrations of the Translations measure. Multiple Regression was used to examine Translations gain scores with respect to SAT-10 and Algebra Composite outcomes (see Tables 29-32) for students in algebra classrooms.

Table 29

Overall Results for Translations Regression Model Predicting SAT-10

Model Summary					
	R	R^2	Adjusted R^2	R^2 change	
	.470	.221	.206	.031	
ANOVA					
Source	*SS*	*df*	*MS*	*F*	*p*
Regression	3206.21	2	1603.11	14.73	.000
Residual	11318.69	104	108.83		
Total	14524.90	106			

Table 30

Regression Coefficients for Translations Model Predicting SAT-10

Variable	*b*	*SE*	*t*	*B*	*sr*	*p*
Intercept	30.68	1.92	15.95			.000
Data point 2	0.53	0.17	3.14	0.32	.27	.002
Data point 5	0.33	0.16	2.02	0.21	.18	.046

Note. *SE* = standard error, *sr* = semipartial correlation.

Table 31

Overall Results for Translations Regression Model Predicting Algebra Composite

Model Summary						
	R	R^2	Adjusted R^2	R^2 change		
	.660	.435	.423	.073		
ANOVA						
Source		*SS*	*df*	*MS*	*F*	*p*
Regression		154.32	2	77.16	35.48	.000
Residual		200.10	92	2.18		
Total		354.42	94			

Table 32

Regression Coefficients for Translations Model Predicting Algebra Composite

Variable	*b*	*SE*	*t*	*B*	*sr*	*p*
Intercept	3.78	0.29	13.00			.000
Data point 2	0.17	0.03	4.53	0.43	.36	.000
Data point 5	0.09	0.03	3.44	0.32	.27	.001

Note. *SE* = standard error, *sr* = semipartial correlation.

How much gain can be expected? Although growth could not be modeled for the Translations measure, mean gain during the fall can be examined. At the first administration of the Translations measure (i.e., data point 2), the mean score was 8.49 items correct (see Table 3). The mean score on the Translations measure at the second administration (i.e., data point 5) was 11.88 items correct (see Table 3). A paired samples t-test was conducted to compare mean Translations scores across the two administrations. Differences in mean scores were significant $t(106) = 5.20$, $p < .001$. Translations scores significantly increased during the fall, with a mean gain of 3.52 items.

Does gain predict outcomes, above and beyond initial skills? Although student performance on the first administration of the Translations measure (i.e., data point 2) explained more variance in scores on the SAT-10 and Algebra Composite, the unique contribution of the second administration of the Translations measure (i.e., data point 5) to outcome scores was significant for the SAT-10 model, $F(2, 104) = 14.73, p < .001$ and the Algebra Composite model, $F(2, 92) = 35.48$, $p < .001$. An examination of the unstandardized betas for the SAT-10 and Algebra Composite models (see Tables 30 and 32) indicates the amount of raw change in outcome measure scores that corresponds with a one-unit change in the predictor. For each additional item scored correct on the Translations measure, it is expected that algebra students will answer between .33 and .53 additional items correct on the SAT-10 and between .17 and .09 additional items correct on the Algebra Composite, depending on time of administration. These results indicate gains in Translations measures explain variance in general math and algebra outcomes. Consequently, the Translations measure demonstrates promise for use as a progress monitoring measure for algebra students in 8th grade. Future studies should examine this measure in greater depth and attempt to model growth using additional data points.

CHAPTER V

DISCUSSION

This study investigated the technical adequacy of mixed computation and algebra CBM in middle school with respect to general math and algebra outcomes. With the intent of contributing to the research base concerning evaluation tools available for use in middle school mathematics classrooms, this study examined the reliability, validity, and expected growth rates of existing CBMs. Moreover, this study targeted progress monitoring in middle school algebra as an important area of development because experts indicate algebra acts as an access point for success in advanced mathematics and beyond and formative assessment is linked to positive results for students. Results from this study indicate that students make growth on mixed computation and algebra CBMs during the fall of 8^{th} grade; however, growth does not appear to predict mid-year general math or algebra outcomes above and beyond initial skills. This chapter discusses these findings in light of current literature, identifies limitations of the present study, and provides suggestions for additional, related research.

Implications for Theory and Practice

Research suggests, to validate a measure for use in progress monitoring, a three-stage process is needed, involving an exploration of (a) the technical adequacy of the measure at one point in time; (b) the technical features associated with slope, perhaps relative to a target domain; and (c) the use data gleaned from the measure for informing instruction and intervention (Fuchs, 2004; Hojnoski et al., 2009). This study explored the technical adequacy of mixed computation and algebra CBM at the first and second stages of the validation process. Correlations observed between mixed computation CBM and outcome measures were slightly lower than those identified in earlier studies (e.g., Thurber et al., 2002), while correlations between algebra CBM and outcome measures were moderately higher than those identified previously (e.g., Foegen et al., 2005). Importantly, although rates of progress did not predict outcome measures above and beyond initial skills, growth for students on mixed computation and algebra CBM in the fall of 8th grade was significant. Slopes of progress for mixed computation and algebra CBM are largely undefined in published research; thus, it is difficult to identify whether the findings of the present study converge with other research. Consequently, it is important that research continues to focus on stage two developments with middle school progress monitoring measures and examine whether the findings in this study are consistent with new research.

Using the standards recommended by Salvia and Ysseldyke (2007), correlations observed in this study indicate acceptable levels of reliability for algebra CBM in 8th grade, but limited reliability for M-CBM2. In addition, although the CBM measures studied may not be technically adequate for important individual educational judgments with the sample assessed in this study, algebra CBM approaches a standard of validity evidence based on relations to the SAT-10 and Algebra Composite that may support its use as a screening measure. In this regard, algebra CBM may be appropriate for lower stakes decisions, such as day-to-day instructional planning.

Results of this study also indicate initial status predicts outcomes across all measures (except for M-CBM2, for predicting the Algebra Composite). Predicted initial status on M-CBM2, Basic Skills, Algebra Foundations, and Translations contributed to increases in SAT-10 and Algebra Composite scores. These contributions were significant, which demonstrates that, for 8th grade students in this sample, initial status on mixed computation and algebra CBM was indicative of later performance on an established measure of general math and algebra skills. Findings about the significance of initial status for predicting mid-year outcomes provides further evidence that algebra CBM may allow classroom teachers to make decisions about whether students are on track at the beginning of fall to meet mid-year goals, and may support instructional decision-making by encouraging teachers to instruct upon skills students cannot demonstrate across the measures.

Finally, this study provides some evidence to support the use of measures with strong face validity. Teachers of pre-algebra and algebra students may be less willing to use M-CBM2 than algebra CBM based on the appearance of the measures, because algebra CBM looks more relevant to the instruction that occurs in pre-algebra and algebra classrooms. Based on the preliminary analyses conducted in this study, teachers' inclinations to prefer algebra CBM over mixed computation CBM in 8th grade math classrooms is likely appropriate: correlations between M-CBM2 and outcome measures were weak, while correlations between algebra CBM and outcome measures were moderate to strong. Also, though initial status on M-CBM2 was predictive of SAT-10 outcomes, it was not predictive of Algebra Composite scores, which suggests, for algebra teachers working to prepare their students for courses in algebra and advanced mathematics, M-CBM2 may have less immediate relevance.

Limitations

Several limitations can be associated with this study. First, sample constraints may have influenced the generalizability of study findings, because the study was limited to 8th grade students and was conducted only in the Pacific Northwest region of the United States. Also, the number of classrooms containing student participants was on the low end of the acceptable range for HLM analyses. As a result, random variation could not be fully explored for each of the measures studied, which may limit understanding of the impact of classrooms on student progress, especially in attempts to quantify any differences in slopes between students in pre-algebra and algebra classrooms or examine instructional differences that my have impacted progress across teachers or schools.

Second, based on initial agreements with teachers in participating classrooms and subsequent requests for additional time for students to complete outcome measures, SAT-10 administration procedures were not identical across all three participating school sites. At schools A and B, students were assessed across two days. On the first day of testing the researcher administered the test; on the second day of testing, the teacher provided additional time for students to finish test items. In contrast, at school C, SAT-10 testing was conducted in one day, by the researcher. In addition, while all students received at least 80 minutes to complete both subtests of the SAT-10, some of the students did not complete the test, which likely affected student scores on the outcome measure, and more heavily impacted the second administered subtest of the SAT-10 (i.e., Math Procedures). By not allowing as much time as was needed to complete the test or constraining time allowed across sites for consistency, variation in results is difficult to analyze.

Third, school schedule changes mid-study affected student participation in the study. At school A, administrative efforts were made to keep students with the same math teacher, despite changes to class composition (e.g., Student 1 moved from period 2 with teacher A to period 5 with teacher A). At school B, student schedules were largely overhauled, especially in pre-algebra classrooms, based on changes to the school's master schedule. As a result, the number of participating students from pre-algebra classrooms at school B is low. In addition, across the three schools, a total of five students moved from pre-algebra to algebra during the study; however these students were retained in the pre-algebra condition for the purposes of analyses. The nature of conducting research in schools requires flexibility, thus, means for managing missing data were planned at the onset of the study and used during analyses to limit the impacts of attrition on results.

Fourth, classroom context was different in each of the schools. Curricula varied in author, content, and year, although all were considered traditional in design. Also, teachers approached study participation differently. Teachers of participating classrooms were provided with raw and summarized student data after each measurement occasion and told they could use the data to support instructional planning. Although all teachers talked with students about the meaning of the assessments and the rationale for their participation, incentives for student improvement varied across schools. To motivate students to continue to put effort into completing assessments after the first two administration sessions, one teacher chose to announce and provide extra credit for students who made progress between measurement occasions. Another teacher decided to not assign homework on days the researcher came to the classroom to administer assessments. A third teacher provided no extrinsic reinforcement for student performance or participation. Although these factors have been documented, sample size prohibits a comprehensive analysis of the effects of these differences across sites.

Future Directions

To address the limitations and further examine the results of this study in additional contexts, research should explore growth rates with an increased number of classrooms, across geographical regions and grade levels. Also, because slopes did not predict general math or algebra outcomes above and beyond initial skills, future studies examining growth rates for these measures should extend the data collection timeline (e.g., to the end of the school year) and/or increase the total number of data points in accordance with recommendations for best practices in methodology (e.g., Raudenbush & Bryk, 2002). In addition, when algebra CBM is used in practice, measure developers recommend that the first data point not be used for interpretation and instead be used as a stabilization point to allow students to become familiar with the tasks of the measures

(Foegen, personal communication, April 21, 2010). If the first administration for Algebra Foundations was removed from the data set, mean scores would demonstrate a consistently positive linear trend (as would the scores for M-CBM2). The first administration was included in analyses to more accurately model student behavior across fall for the purposes of research. However, it may be useful to examine study results excluding the first data point to explore any differences in growth on the measures or in the predictive relation between initial skills or slope and SAT-10 or Algebra Composite scores.

If the CBM measures studied are to be used at a single point in time as an indicator of mid-year general math or algebra outcomes, it will be important to examine the use of the measures for screening. Technically adequate screening assessments allow educators to make decisions about student skill in a broad mathematical domain, and study results suggest the measures studied are indicative of broad general math and algebra skills. In this regard, it may be valuable to examine how measures might be used as a set (e.g., administer two or three measures and take the mean or median) to predict math outcomes. However, in order to validate any of the studied CBMs for screening, additional research is needed with larger samples and in the context of instructional decision making to examine the technical adequacy and utility of the measures for this purpose. Also, reading has been identified as an area of development that may be highly correlated with math skill (e.g., Fuchs et al., 2004; Thurber et al., 2002). It may be important to examine the role of oral reading fluency with respect to student growth and outcomes in mathematics.

Although measure reliability and validity for a purpose are necessary features of any assessment, research suggests that additional components of technical adequacy should be established for a measure to be appropriate for progress monitoring (Ardoin & Christ, 2009; Francis et al., 2008). Because progress monitoring measures are intended to provide a linear snapshot of student skill in a content domain, it is important that any variation in student scores over time can be attributed to student growth rather than standard error (Francis et al., 2008). Given this premise, technically adequate progress monitoring measures require an appropriate number of alternate forms and documented consistency in the difficulty across forms (Ardoin & Christ, 2009). For the Translations measure, the development of alternate forms is of particular importance, if student growth on the measure is to be studied.

Similarly, although measure content appears to target algebra skills, future studies should examine item level content of mixed computation and algebra CBM relative to well-respected outcome measures, such as the SAT-10, to explicitly identify the standards measured by each test. By moving research on middle school math CBM into these domains and linking measures to algebra standards, results have the potential to better inform uses of the measures for data based decision-making in mathematics. This research did not purport to explore these features of mixed computation or algebra CBM;

thus, to examine the item and form difficulty, internal consistency, and specific skills targeted by the measures studied, additional analysis of test items and forms is needed.

In addition, this study did not examine content alignment between classroom instruction and outcome measures. If the SAT-10 or Algebra Composite do not accurately represent the content taught in the pre-algebra and algebra classrooms studied, the finding that slope does not predict these outcomes above and beyond initial skills has little meaning: It is possible that an outcome measure that may be more aligned with either the CBMs administered or the content taught to students may result in different findings. To gain additional insight on this issue, future studies should explore the alignment between content taught to student participants, the content assessed by the outcome measure, and the content assessed by each CBM administered. For example, future studies with these measures might employ the use of an algebra-focused outcome measure derived from algebra course final exams or, at the least, utilize a teacher rating system to identify the extent to which an outcome measure and administered CBMs are aligned with intended course content.

Recent research also indicates that instructional variables can impact growth patterns across classrooms (Hojnoski et al., 2008). In this regard, it's possible that evaluating the technical adequacy of an assessment using outcome measures administered months after the initial administration of the tool being evaluated (as was done in the present study) may not provide a sufficient depiction of the technical adequacy of the tool or the nature of skill development. Because instruction invariably occurs between administration of the initial measure and administration of the outcome measure for students in schools, it is difficult to define whether changes in student performance are the result of accurate measurement (i.e, true changes in student skill), variability in measure items or forms, or a combination of factors. If measure difficulty has not been studied, the extent to which performance on one measure predicts skills on a second, distal measure becomes even more ambiguous. In this regard, in addition to analyzing measure difficulty and constructs assessed, it is important to document the quality of instruction students receive when measures are studied with respect to later outcomes, especially if differences emerge between classrooms, instructors, or schools (Raudenbush & Bryk, 2002). By conducting observations throughout the assessment period (e.g., tracking the use of other evidence-based practices in mathematics, documenting student engagement and other alterable classroom variables such as student practice), changes in student skill can be more accurately matched to measure sensitivity across student groups.

Conclusion

Assessment—the process of collecting data to make decisions about students—is important because results allow educators to better identify student strengths and weaknesses and make decisions about instruction with increased accuracy (Salvia & Ysseldyke, 2007). In the context of expert recommendations, assessment in mathematics needs additional study. Specifically, experts call for development and study of objective measures of mathematics achievement that provide direct information to teachers and students about the learning process (NCTM, 2000). Simultaneously, experts recommend an increased focus on algebra content in math instruction and assessments (NMAP, 2008), and research on evidence-based practices in mathematics indicates formative assessment has the power to increase student outcomes when results are used by teachers and communicated to students (Gersten et al., 2008). Consequently, studies of formative assessment measures that are technically adequate for instructional decision-making in algebra are needed to support student achievement in mathematics.

In consideration of the technical features described by Francis et al. (2008), which are required for a measure to be appropriate for progress monitoring (i.e., administration across regular intervals, brief and easy to administer, scores use a consistent metric, scores are predictive of important end of year outcomes, and forms are free from measurement artifacts), this study provides evidence that existing mixed computation and algebra CBM meet the first three criteria. Results also indicate initial status is predictive of mid-year general math performance, which provides some evidence for the fourth criterion. Criterion five was not addressed in this study. With respect to the criteria set forth by the NCPM (2007), mixed computation and algebra CBM probes appear to be reliable, sensitive to student improvement, and have specified rates of improvement; however, more research is needed to identify whether the tools are valid for progress monitoring, determine if they can be linked to improved student learning or teacher planning, and specify adequate yearly progress benchmarks.

In summary, despite the indication that the measures studied may have questionable technical adequacy for progress monitoring with this sample, the measures may prove useful for other means of formative evaluation. Given the limitations of any single study, progress monitoring should not be ruled out for any of the measures studied. Instead, more research is needed to identify whether student growth is stable enough on the studied CBMs to be used for progress monitoring.

APPENDIX A

MATH COMPUTATION CURRICULUM BASED MEASUREMENT, 2^{ND} VERSION

AIMSweb® Mathematics Computation (Rel. 2006) Progress Monitor #2 - Grade 8

You have 4 minutes to write your answers to several kinds of math problems. Look at each problem carefully. Try to work each problem, but if you REALLY don't know how to do it, put an X over it and go to the next one. Don't skip around. Some problems require you to read the instructions on the page. Reduce fractions to their most common form, and round decimals to the thousandths place.

Student Name: ____________________ **Grade:** __________ **Teacher Name:** ____________________

$1 / \frac{9}{10} =$	Convert to Decimal $\frac{1}{4} =$	90% of 74 =	Convert to Fraction .4 =	$\frac{2}{3} + \frac{7}{9} =$
Convert to Fraction .5 =	$\begin{array}{r} 20.81 \\ + 17.59 \\ \hline \end{array}$	$\begin{array}{r} 35.09 \\ - 2.54 \\ \hline \end{array}$	Convert to Fraction .75 =	$\begin{array}{r} 57.9 \\ \times 4 \\ \hline \end{array}$
$\frac{8}{9} \cdot \frac{8}{9} =$	$5.4\overline{)86.0}$	$20\overline{)110.0}$	$\begin{array}{r} 33.4 \\ \times 5 \\ \hline \end{array}$	$\begin{array}{r} 59.6 \\ \times 1.2 \\ \hline \end{array}$
Convert to Decimal $\frac{4}{5} =$	$\frac{2}{5} / \frac{8}{9} =$	$\begin{array}{r} 39.13 \\ - 8.02 \\ \hline \end{array}$	Convert to Decimal $\frac{1}{8} =$	Convert to Fraction .9 =
$78\overline{)101}$	$14\overline{)864}$	$3.6\overline{)24.2}$	$\begin{array}{r} 47.76 \\ + 2.76 \\ \hline \end{array}$	$\begin{array}{r} 66.95 \\ + 28.23 \\ \hline \end{array}$
$\begin{array}{r} 25.5 \\ \times 3 \\ \hline \end{array}$	Convert to Fraction .1 =	$\begin{array}{r} 20.3 \\ \times 2.8 \\ \hline \end{array}$	20% of 20 =	$\frac{8}{9} + \frac{5}{9} =$

AIMSweb® Mathematics Computation (Rel. 2006) Progress Monitor #2 - Grade 8

You have 4 minutes to write your answers to several kinds of math problems. Look at each problem carefully. Try to work each problem, but if you REALLY don't know how to do it, put an X over it and go to the next one. Don't skip around. Some problems require you to read the instructions on the page. Reduce fractions to their most common form, and round decimals to the thousandths place.

Student Name: __________________ **Grade:** ________ **Teacher Name:** __________________

$6.5\overline{)39.9}$	Convert to Fraction .2 =	Convert to Fraction .8 =	$4.5\overline{)66.1}$	Convert to Fraction .25 =
8×7.5	Convert to Decimal $\frac{9}{10}$ =	75% of 16 =	94.9×3.9	$34.89 + 13.46$
Convert to Decimal $\frac{7}{10}$ =	90% of 55 =	10×2	75% of 24 =	20.6×3
$94.85 + 12.98$	Convert to Fraction .6 =	$81\overline{)150}$	$8\overline{)64.0}$	$\frac{2}{7} \times \frac{8}{9} =$
11×6	$\frac{6}{7} + \frac{5}{7} =$	976×30	$\frac{5}{9} + \frac{1}{3} =$	$11\overline{)91.3}$
$11\overline{)103.4}$	149×73	$31.55 - 9.09$	36.2×6.1	$1 - \frac{8}{9} =$

APPENDIX B

BASIC SKILLS

Solve: $9 + a = 15$	$a =$
Evaluate: $12 + (-8) + 3$	
Simplify: $2x + 4 + 3x + 5$	
Solve: $12 - e = 4$	$e =$
Simplify: $4(3 + s) - 7$	
Simplify: $b + b + 2b$	
Solve: $\frac{r}{6} = \frac{12}{18}$	$r =$
Simplify: $7 - 3(f - 2)$	
Evaluate: $-5 + (-4) - 1$	
Solve: $63 + c = 9$	$c =$
Simplify: $2(s - 1) + 4 + 5s$	
Simplify: $8m - 9(m + 2)$	
Solve: 3 ft. = 1 yd. ____ ft. = 9 yds.	
Evaluate: $4 - (-2) + 8$	
Simplify: $2k + 3 - 5(k + 7)$	

Solve: $10 - 6 = g$	$g =$
Simplify: $9 - 4d + 2 + 7d$	
Simplify: $5(b - 3) - b$	
Solve: $q \cdot 5 = 30$	$q =$
Evaluate: $8 - (-6) - 4$	
Simplify: $2 + w(w - 5)$	
Solve: 1 ft. = 12 in. 5 ft. = ____ in.	
Simplify: $4 - 7b + 5(b - 1)$	
Simplify: $s + 2s - 4s$	
Solve: $x + 4 = 7$	$x =$
Simplify: $-5(q + 3) + 9$	
Evaluate: $9 + (-3) - 8$	
Solve: $\frac{12}{2} = \frac{48}{e}$	$e =$
Simplify: $y^2 + y - 4y + 3y^2$	
Simplify: $3(c + 2) - 2c$	

OSEP Award# H324C030060

Solve: $3 \cdot 8 = m$	$m =$
Evaluate: $-9+5+8$	
Simplify: $x+2(x-5)-3$	
Solve: $d-5=4$	$d=$
Simplify: $5(3+f)-2f+6$	
Simplify: $5-2b+4(b+3)$	
Solve: 4 qts. = 1 gal. _____ qts. = 3 ¼ gals.	
Simplify: $4(y+1)-8y$	
Evaluate: $14-7+(-3)$	
Solve: $\frac{36}{6}=s$	$s=$
Simplify: $-3w^2+5w^2-5+12$	
Simplify: $9-4(v+2)$	
Solve: $4r=28$	$r=$
Simplify: $16+2(t-4)-3t$	
Simplify: $c-3(c+2)+8$	

Solve: $\frac{1.5}{3}=\frac{h}{9}$	$h=$
Simplify: $7b-4-3-2b$	
Simplify: $2e-3(e-4)$	
Solve: $6+7=y$	$y=$
Evaluate: $-5+6-6$	
Simplify: $4+10(1-r)$	
Solve: 2.5 cm. = 1 in. _____ cm. = 6 in.	
Simplify: $6a+2a-9+3a^2$	
Evaluate: $-1+4+(-7)$	
Solve: $\frac{500}{j}=\frac{10}{2}$	$j=$
Simplify: $-3(u+3)-2u+5$	
Simplify: $2c-3c-c$	
Solve: $h+6=8$	$h=$
Evaluate: $-2+(-5)+(-8)$	
Simplify: $3z-8z+2+9$	

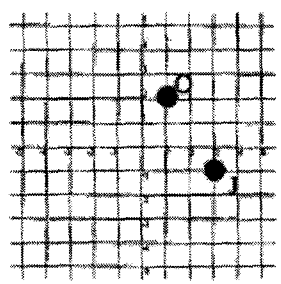
OSEP Award# H324C030060

APPENDIX C

ALGEBRA FOUNDATIONS

Find the ordered pair for each point:

J(,) O(,)



Fill in the empty box:

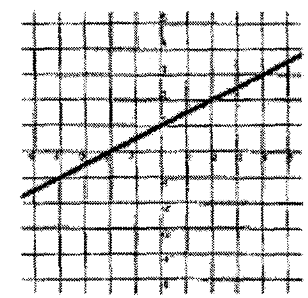
s	$3s$
6	18
7	21
8	
9	27

Fill in the empty box:

n	$4n+7$
-1	3
-2	
-3	-5
-4	-9

Fill in the empty box:

b	
-2	-5
0	-3
3	0
5	2



What is the slope?

What is the y-intercept?

If $y > 9$, two possible values for y are ____ and _____

Evaluate:
$9 \bullet 4 - 6$

Simplify:
$7f + (2f + f)$

Solve:
$n + 3 = 8$
$n =$

Evaluate $4b + 2$ when

$b = 1$ ________

$b = 3$ ________

Write the expression for this phrase:
6 less than a number

Evaluate:
$(-2) \bullet (-4)$

Graph the expression $m > -5$

Write a word phrase for this expression:
$n + 9$

Evaluate:
$4 + (9 + 3) - 2^2$

Evaluate:
$(-2)^3$

Write the expression for this phrase:
9 multiplied by a number

Evaluate $2x + 4y$ when
$x = 2$ and $y = -3$

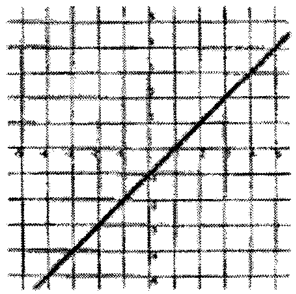
Write a word phrase for this expression:
$10b - 7$

Evaluate $8g - 4$ when

$g = 2$ ________

$g = -2$ ________

Simplify:
$6 - 2(b - 4)$



What is the slope?

What is the y-intercept?

Fill in the empty box:

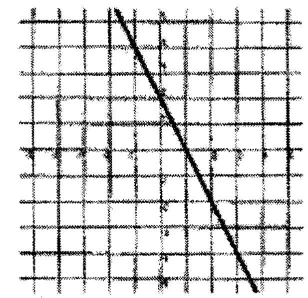
n	
6	4
9	6
12	8
15	10

Fill in the empty box:

t	$2t - 7$
-2	-11
2	-3
6	
10	13

Fill in the empty box:

h	
-5	2
1	8
5	12
10	17



What is the slope?

What is the y-intercept?

Write a word phrase for this expression:
$x + 4$

Evaluate:
$(-12 + 4) + 5$

Write the expression for this phrase:
8 more than twice a number

Solve:
$15 - 8 = x$
$x =$

Evaluate:
4^2

Graph the expression $p \leq 3$

Simplify:
$9x - 3 + 4(x + 9)$

Write the expression for this phrase:
10 divided by a number

If $2a + 4 < 20$, two possible values for a are ________ and ________

Solve:
$24 + x = 6$
$x =$

Evaluate:
$10 - 3 + 8 + 2$

Simplify:
$12n - 5 + 3 - 7n$

Write a word phrase for this expression:
$h \cdot 5$

Evaluate:
$(-3)(9 - 7)$

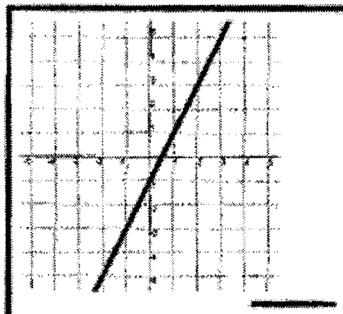
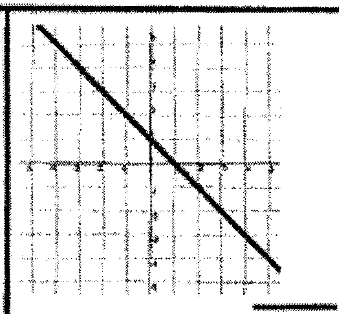
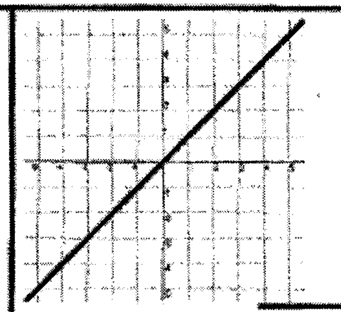
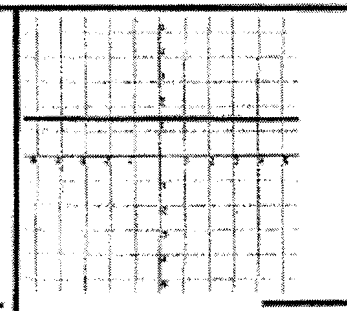
Evaluate:
$\sqrt{81}$

Solve:
$6t = 36$
$t =$

APPENDIX D

TRANSLATIONS

A	B	C	D
$y = x$	$y = 2x - 1$	$y = 1.5$	$y = -x + 1$

x	y
2	1.5
1	1.5
0	1.5
-1	1.5
-2	1.5

x	y
2	-1
1	0
0	1
-1	2
-2	3

x	y
2	3
1	1
0	-1
-1	-3
-2	-5

x	y
4	4
2	2
0	0
-2	-2
-4	-4

x	y
4	-3
2	-1
0	1
-2	3
-4	5

Mark needs to find half the width of pieces of pipe he is cutting to make a soccer goal. The width of the pipe is 3 inches. He wrote this equation to show the relationship between the length and the width of the pieces he will cut. ____

Every day that Cindy waters the garden, she earns a dollar. She wrote this equation to show the relationship between the number of days she waters the garden and the number of dollars she will earn. ____

Joe has one dollar in his wallet. He wrote this equation to show the relationship between the number of dollars he borrows from his friends for lunch and the total amount of money he has or owes. ____

The class earns \$2 for each magazine subscription sold in the fund-raiser. A \$1 fee per student is charged for a processing fee. Cindy wrote this equation to show the relationship between the number of magazines sold and the profit. ____

The flood waters are receding at a rate of 1 foot per day. The river is currently at 1 foot above flood stage. Tom wrote this equation to show the relationship between the number of days and the height of the river compared to flood stage. ____

OSEP Award# H324C030060

A

x	y
4	16
2	4
0	0
-2	4
-4	16

B

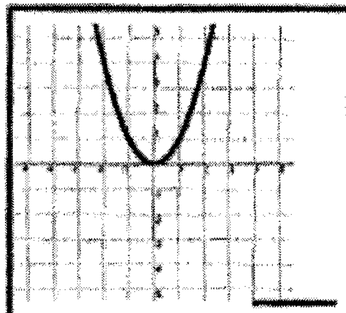
x	y
2	4
1	2
0	1
-1	$\frac{1}{2}$
-2	$\frac{1}{4}$

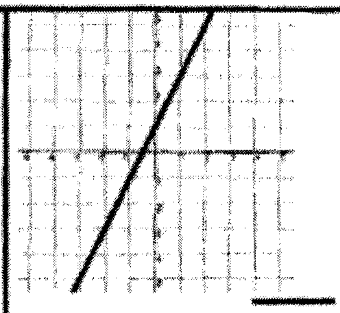
C

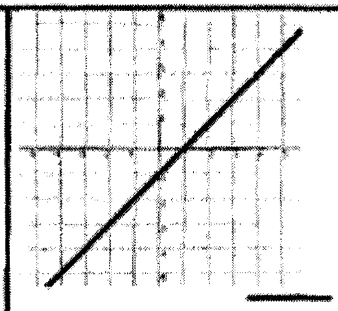
x	y
2	5
1	3
0	1
-1	-1
-2	-3

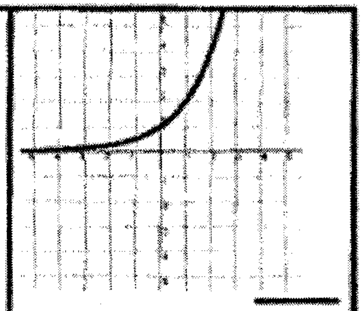
D

x	y
4	3
2	1
0	-1
-2	-3
-4	-5

 ____

 ____

 ____

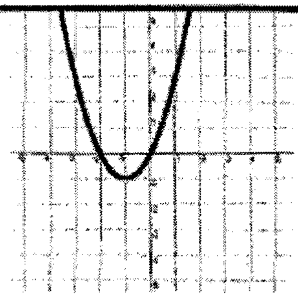
 ____

$y = 2x + 1$ ____	$y = 2^x$ ____	$y = x - 1$ ____	$y = x^2$ ____	$3y = 6x + 3$ ____

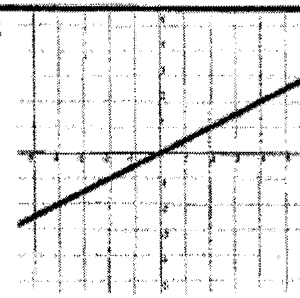
Mr. Jones is going to give a true/false test. He made this data table to show the number of possible answer combinations his students can give on the test.	____
Sue made this data table to figure out how many inches of wire she needs for a bracelet. Each bracelet uses two strands and she needs to add an extra inch to make a hook to fasten the bracelet.	____
Sam's allowance changes every year. Each month his mom pays him a dollar for each year he has lived, multiplied by his age. Sam made this data table to figure out his allowance.	____
Every time Hans delivers newspapers, he keeps one for his family. Hans made this data table to show how many newspapers he delivers to families on his route.	____
Tim's washing machine 'eats' socks. The first time he lost one sock in the wash. Now, every time he washes a load of clothes, he loses two socks. Tim made this data table to figure out how many socks he has lost.	____

USEF Award# HS24C030060

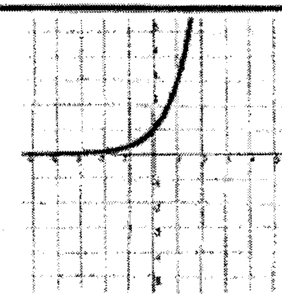
A



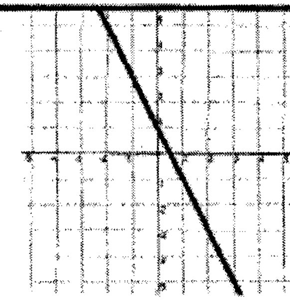
B



C



D



x	y
2	-3
1	-1
0	1
-1	3
-2	5

x	y
4	2
2	1
0	0
-2	-1
-4	-2

x	y
2	9
1	3
0	1
-1	$\frac{1}{3}$
-2	$\frac{1}{9}$

x	y
2	8
1	3
0	0
-1	-1
-2	0

$y = -2x + 1$ ____	$y = x^2 + 2x$ ____	$y = 3^x$ ____	$y = x(x + 2)$ ____	$y = \frac{1}{2}x$ ____

Matt built a maze for his gerbil. Each time the gerbil comes to an intersection, it can go three possible ways. Matt made this graph to show the total possible number of routes for the gerbil through the maze.	____
LaShaya's mom makes her save half of what she earns in the summer for college. She made this graph to show how much money she will earn for her college fund this summer.	____
A diving board is one foot above the surface of the pool. An average diver drops twice his height when he steps off the board. Marcus made this graph to show the diver's depth in the water.	____
Ming Hui has two cats, Oscar and Otis. She knows that Oscar eats twice as much as Otis. She made this graph to show how much Otis eats.	____
Tammy is making a backdrop for the school play. She needs to add on to a square piece of wood. The piece she will add is the same height as the square, but only 2 feet wide. Tammy made this graph to show the area of the backdrop.	____

OSEP Award# H324C030060

REFERENCES

Allinder, R. M., Bolling, R., Oats, R., & Gagnon, W. A. (2000). Effects of teacher self-monitoring on implementation of curriculum-based measurement and mathematics computation achievement of students with disabilities. *Remedial and Special Education, 21*, 219–226.

American Educational Research Association, American Psychological Association, & the National Council on Measurement in Education (1999). *Standards for educational and psychological testing.* Washington, DC: AERA.

Ardoin, S. P., & Christ, T. J. (2009). Curriculum-based measurement of oral reading: Standard errors associated with progress monitoring outcomes form DIBELS, AIMSweb, and an experimental passage set. *School Psychology Review, 38*(2), 266-283.

Ball, D., & Bass H. (2002). Toward a practice-based theory of mathematical knowledge for teaching. Plenary Lecture at the 26th Annual Meeting of the Canadian Mathematics Education Study Group, Kingston, Ontario, Canada.

Burns, M. K., Deno, S. L., & Jimerson, S. R. (2007). Toward a unified response-to-intervention model. In S. R. Jimerson, M. K. Burns, & A.M. VanDerHeyden (Eds.), *Handbook of response to intervention* (pp. 428-439). New York: Springer.

Burns, M. K., VanDerHeyden, A. M., & Jiban, C. (2006). Assessing the instructional level for mathematics: A comparison of methods. *School Psychology Review, 35*, 401-418.

Calhoon, M. B. (2008). Curriculum-based measurement for mathematics at the high school level: What we do not know…what we need to know. *Assessment for Effective Intervention, 33,* 234-239.

Calhoon, M. B., and Fuchs, L. S. (2003). The effects of peer-assisted learning strategies and curriculum-based measurement on the mathematics performance of secondary students with disabilities. *Remedial and Special Education, 24*(4), 235-245.

Chard, D. J., Ketterlin-Geller, L. R., Jungjohann, K., & Baker, S. K. (2009). Evidence-based math instruction: Developing and implementing math programs a the core, supplemental, and intervention levels. In G. G. Peacock, R. A. Gimpel, & E. J. Daly (Eds.), *Practical handbook of school psychology: Effective practices for the 21st century* (pp. 287-299). New York: Guilford.

Checkley, K. (2001, October). Algebra and activism: Removing the shackles of low expectations: A conversation with Robert P. Moses. *Educational Leadership, 59*, 6-11.

Clarke, B., & Shinn, M. R. (2004). A preliminary investigation into the identification and development of early mathematics curriculum-based measurement. *School Psychology Review, 33*, 234-248.

Deno, S. L. (2002). Problem solving as best practice. In A. Thomas and J. Grimes (Eds.), *Best practices in school psychology* (4th ed., pp. 37-56). Bethesda, MD: National Association of School Psychologists.

Deno, S. L., Fuchs, L. S., Marston, D., & Shin, J. (2001). Using curriculum-based measurement to establish growth standards for students with learning disabilities. *School Psychology Review, 30*, 507-524.

Education Commission of the States (1997, October 20). Mathematics equals opportunity. White paper prepared for Secretary of Education Richard W. Riley. Available from http://www2.ed.gov/pubs/math/index.html

Foegen, A. (2000). Technical adequacy of general outcomes measures for middle school mathematics. *Assessment for Effective Intervention, 25*, 175-203.

Foegen, A. & Deno, S. L. (2001). Identifying growth indicators for low-achieving students in middle school mathematics. *The Journal of Special Education, 35*, 4-16.

Foegen, A. (2006). Evaluating instructional effectiveness. In M. Montague & A. K. Jitendra (Eds.), *Teaching mathematics to middle school students with learning disabilities* (pp. 108-132). New York: Guilford.

Foegen, A. (2008). Progress monitoring in middle school mathematics: Options and issues. *Remedial and Special Education, 29*, 195-207.

Foegen, A. (2009, April 4). Putting Algebra Progress Monitoring Into Practice: An Illustration From the Field. Paper presented at the Council for Exceptional Children Convention and Expo, Seattle, WA.

Foegen, A., Jiban, C., & Deno, S. (2007). Progress monitoring measures in mathematics: A review of the literature. *The Journal of Special Education, 41*, 121-139.

Foegen, A., Olson, J. R., & Perkmen, S. (2005). Reliability and criterion validity of five algebra measures in Districts B and C. (Technical Report 7). Project AAIMS, Department of Curriculum and Instruction, Iowa State University, Ames, IA.

Francis, D. J., Santi, K. L., Barr, C., Fletcher, J. M., Varisco, A., & Foorman, B. R. (2008). Form effects on the estimation of students' oral reading fluency using DIBELS. *Journal of School Psychology, 46*, 315 – 342.

Fuchs, L. S. (2004). The past, present, and future of curriculum-based measurement research. *School Psychology Review, 33*, 188–192.

Fuchs, L. S., & Fuchs, D. (1990). The role of skills analysis in curriculum-based measurement in math. *School Psychology Review, 19*(1), 6-23.

Fuchs, L. S., Fuchs, D. Hamlett, C. L., Walz, L., & Germann, G. (1993). Formative evaluation of academic progress: How much growth can we expect? *School Psychology Review, 22*(1), 27-48.

Fuchs, L. S., Fuchs, D., Phillips, N. B., Hamlett, C. L., & Karns, K. (1995). Acquisition and transfer effects of classwide peer-assisted learning strategies in mathematics for students with varying learning histories. *School Psychology Review, 24*, 604–620.

Fuchs, L. S., Fuchs, D., & Prentice, K. (2004). Responsiveness to mathematical problem-solving instruction: Comparing students at risk of mathematics disability with and without risk of reading disability. *Journal of Learning Disabilities, 37*, 293–306.

Fuchs, L. S., Fuchs, D., Prentice, K., Burch, M., Hamlett, C. L., Owen, R., …Schroeter, K. (2003). Enhancing third-grade students' mathematical problem solving with self-regulated learning strategies. *Journal of Educational Psychology, 95*, 306-315.

Fuchs, L. S., & Shinn, M. R. (1989). Writing CBM IEP objectives. In M. R. Shinn (Ed.), *Curriculum-based measurement: Assessing special children* (pp. 130-152). New York: Guilford.

Gersten, R., Baker, S., & Chard, D. (2006, November 13). Effective instructional practices for students with difficulties in mathematics. Presented at the Center on Instruction Math Summit, Annapolis, MD.

Gersten, R., Beckmann, S., Clarke, B., Foegen, A., March, L., Star, J. R., …Witzel, B. (2009). *Assisting students struggling with mathematics: Response to Intervention (RtI) for elementary and middle schools* (Practice Guide Report No. NCEE 2009-4060). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, US Department of Education.

Gersten, R. & Chard, D. (1999). Number sense: Rethinking arithmetic instruction for students with mathematical disabilities. *The Journal of Special Education, 33*, 18-28.

Gersten, R., Chard, D. J., Jayanthi, M., Baker, S. K., Morphy, P. & Flojo, J. (2008). *Mathematics instruction for students with learning disabilities or difficulty learning mathematics: A synthesis of the intervention research.* Retrieved from the Center on Instruction website: http://www.centeroninstruction.org/ resources.cfm?category=math

Gersten, R., Jordan, N. C., & Flojo, J. R. (2005). Early identification and interventions for students with mathematics difficulties. *Journal of Learning Disabilities, 38*(4), 293-304.

Good III, R. H. & Jefferson, G. (1998). Contemporary perspectives on curriculum-based measurement validity. In M. R. Shinn (Ed.), *Advanced applications of curriculum-based measurement* (pp. 61-88). New York: Guilford.

Helwig, R., Anderson, L., & Tindal, G. (2002). Using a concept-grounded curriculum-based measure in mathematics to predict statewide test scores for middle school students with LD. *The Journal of Special Education, 36*, 102-112.

Hintze, J. M. (2008). Conceptual and empirical issues of RTI. Unpublished manuscript, University of Massachusetts at Amherst.

Hojnoski, R. L., Silberglitt, B., Floyd, R. G. (2009). Sensitivity to growth over time of the preschool numeracy indicators with a sample of preschoolers in Head Start. *School Psychology Review, 38*(3), 402-418.

Individuals with Disabilities Education Improvement Act of 2004. Pub. L. No. 108-446, 118 Stat. 2647 (2005).

Jitendra, A. K., Hoff, K., Beck, M. M. (1999). Teaching middle school students with learning disabilities to solve word problems using a schema-based approach. *Remedial and Special Education, 20*, 50-64.

Jitendra, A., DiPipi, C. M., Perron-Jones, N. (2002). An exploratory study of schema-based word-problem-solving instruction for middle school students with learning disabilities: An emphasis on conceptual and procedural understanding. *The Journal of Special Education, 36*, 23-38.

Kame'enui, E. J., Fuchs, L., Francis, D. J., Good, R., O'Connor, R. E., Simmons, D. C., ...Torgeson, J. K. (2006). The adequacy of tools for assessing reading competence: A framework and review. *Educational Researcher, 35*(4), 3-11.

Kame'enui & Simmons (1990). Designing instructional strategies: The prevention of academic learning problems. Columbus, OH: Merrill.

Kane, M. T. (1992). An argument-based approach to validity. *Psychological Bulletin, 112*(3), 527-535.

Kelley, B., Hosp, J. L., & Howell, K. M. (2008). Curriculum-based evaluation and math. *Assessment for Effective Intervention, 33*, 250-256.

Ketterlin-Geller, L. R., Baker, S. K., & Chard, D. J. (2008). Best practices in mathematics instruction and assessment in secondary settings. In A. Thomas and J. Grimes (Eds.), *Best practices in school psychology* (5^{th} ed., pp. 465-475). Bethesda, MD: National Association of School Psychologists.

Lembke, E. S. & Stecker, P. M. (2007). Curriculum-based measurement in mathematics: An evidence-based formative assessment procedure. Portsmouth, NH: RMC Research Corporation, Center on Instruction.

Messick, S. (1986). The once and future issues of validity: Assessing the meaning and consequences of measurement (Research Rep.). Princeton, NJ: Educational Testing Service.

Messick (1989). Meaning and values in test validation: The science and ethics of assessment. *Educational Researcher, 18*, 5-11.

Milgram, R. J. (2005). *The mathematics pre-service teachers need to know.* Stanford, CA: Author.

Miller, S. P., & Hudson, P. J. (2007). Using evidence-based practices to build mathematics competence related to conceptual, procedural, and declarative knowledge. *Learning Disabilities Practice, 22*, 47-57.

National Assessment of Educational Progress (2003-2009). NAEP questions tool. Retrieved from http://nces.ed.gov/nationsreportcard/itmrlsx/search.aspx?subject=mathematics

National Assessment of Educational Progress (2009). *The nation's report card.* Retrieved from http://nationsreportcard.gov/math_2009/

National Center for Education Statistics (2009). *The 2007 Trends In International Mathematics and Science Study.* Retrieved from http://nces.ed.gov/timss/

National Center on Progress Monitoring (2007, December). *Review of progress monitoring tools.* Retrieved from http://www.studentprogress.org/chart/chart.asp

National Center on Progress Monitoring (2010). What is progress monitoring? Retrieved from http://www.studentprogress.org/

National Council of Teachers of Mathematics. (2000). *Principles and standards for school mathematics.* Reston, VA: Author.

National Mathematics Advisory Panel (2008). *The final report of the NationalMathematics Advisory Panel.* Retrieved from http://www.ed.gov/about/bdscomm/list/mathpanel/index. html

National Research Council. (2001). Adding it up: Helping children learn mathematics. Available from the National Academy Press, 500 Fifth Street NW Lockbox 285, Washington, D.C.

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

Pearson Education (2009). Stanford Achievement Series, Tenth Edition. Retrieved from http://www.pearsonassessments.com/HAIWEB/Cultures/en-us/Productdetail.htm?Pid =SAT10C

Salvia, J., & Ysseldyke, J. E. (2007). *Assessment* (10th ed.). Boston: Houghton Mifflin Company.

Schmidt, W. H., Tatto, M. T., Bankov, K., Blomeke, S., Cedillo, T., Cogan, L., …Schwille, J. (2007). The preparation gap: Teacher education for middle school mathematics in six countries (Research Report No. MT21). Retrieved from MSU Center for Research in Mathematics and Science Education website: http://usteds.msu.edu

Shinn, M. R. (2004). Administration and scoring of Mathematics Computation Curriculum-Based Measurement (M-CBM) and math fact probes for use with AIMSweb. Available from www. aimsweb.com

Shinn, M. R., & Bamonto, S. (1998). Advanced applications of curriculum-based measurement: "Big ideas" and avoiding confusion. In M. R. Shinn (Ed.), *Advanced applications of curriculum-based measurement* (pp. 1-31). New York: Guilford.

Stewart, I. (2007). Euler's revolution. *New Scientist, 193*(2596), 48-51.

Stiggins, R. (2001). The unfulfilled promise of classroom assessment. *Educational Measurement: Issues and Practice, 20*(3), 5–15.

Thurber, R. S., Shinn, M. R., & Smolkowski, K. (2002). What is measured in mathematics tests? Construct validity of curriculum-based mathematics measures. *School Psychology Review, 31*(4), 493-513.

VanDerHeyden, A. M., & Burns, M. K. (2005). Using curriculum-based assessment and curriculum-based measurement to guide elementary mathematics instruction: Effect on individual and group accountability scores. *Assessment for Effective Intervention, 30*(3), 15-31.

Wu, H. (1999). Basic skills versus conceptual understanding: A bogus dichotomy in mathematics education. *American Educator, 23*(3), 14–19, 50–52.

Wu, H. (2009, February 21). From arithmetic to algebra. Professional development session presented at the Mathematicians Workshop Series, Eugene, OR.

Wu, H. (2009). What's sophisticated about elementary mathematics? Plenty—that's why elementary schools need math teachers. *American Educator, 33*(3), 4-14.