2011

Exploring content validity, item level analysis and predictive validity for two algebra progress monitoring measures

Subhalakshmi Singamaneni
Iowa State University

Exploring content validity, item level analysis and predictive validity for two algebra progress monitoring measures

by

Subhalakshmi Singamaneni

A thesis submitted to the graduate faculty

in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

Major: Education (Special Education)

Program of Study Committee:
Anne Foegen, Major Professor
Geoffrey Abelson
Gary D. Phye

Iowa State University

Ames, Iowa

2011

Table of Contents

LIST OF FIGURES

LIST OF TABLES

Abstract

This study examines the content validity, item level analysis and predictive validity of two algebra progress monitoring measures. The content in two algebra progress monitoring measures was examined to determine alignment with the Common Core State Standards (CCSS) for algebra. The content for one algebra measure, namely, Algebra Content Analysis (ACA), aligned well with the Common Core State Standards, with each skill tested in the measure aligning with at least one or more CCSS high school algebra domains. In the second algebra measure, Algebra Basic Skills (ABS), three of the five skills tested in the measure aligned with at least one CCSS high school algebra domain, however the remaining two did not align with any of the CCSS high school algebra domains. For item level analyses, item difficulty and discrimination were examined for both algebra measures. Data for analyses were collected from two school districts (A and B). Eighty-three students from District A and fifty-one students from District B participated in this study. Results indicated that the items in both ABS and ACA are mostly found in the 'average difficulty' level ranging from .3 to .9. Items in ACA displayed good discriminating power in terms of student ability in algebra. Item discrimination analysis for ABS was not performed due to an inadequate sample of attempted items. Predictive validity at the subskill category levels for both the measures was examined by correlating scores on the subskills with scores for ITED and ITBS. In ACA, with the exception of the subskill 3.1 – Solve Linear Equation, the remaining subskills' totals did not show encouraging predictive validity with ITBS. ACA 1 and ACA 3 showed weak relations with ITED. With regards to predictive validity of ABS, four subskill categories had moderate relationships with ITED Computation scores. Implications for practice and future research are discussed.

Chapter 1: Introduction

How does one know that teaching is effective? How does one find out that learning is taking place? How do teachers know that their students are making progress? How do schools report the progress of their students? All of these questions can be answered using a set of words that includes tests, assessment, evaluation, and measurement. Curriculum Based Measurement (CBM) (Deno, 1985) is one such kind of testing that monitors student progress and is synonymous with the term ‘progress monitoring’.

Initially developed to test the efficacy of a special education intervention model called data-based program modification (Deno & Mirkin, 1977), the use of CBM has been extended to monitor student progress in general education as well. In addition to monitoring academic progress in students, CBM is also used to screen and identify students at risk, to predict student performance on high stakes tests and develop school wide accountability systems, to measure growth in early childhood, to assess content area learning, and to evaluate literacy skills in students who are hearing impaired and English language learners. The research to develop CBM was spearheaded by Dr. Stanley L. Deno at the University of Minnesota beginning in the 1970s (Deno & Mirkin, 1977).

Earlier, before CBM, teachers depended on commercially developed achievement tests that were standardized and norm referenced (Deno, 1985). These tests are administered annually and provide information about the academic standing of the student at one point in time. What teachers needed (and these tests could not provide) was information about students’ performance that would indicate whether the student was benefiting from instruction and making adequate progress. In other words, commercial achievement tests were of limited use for making instructional decisions (Salmon-Cox, as cited by Deno, 1985). Moreover, research indicated that these standardized norm-referenced tests were technically inadequate for making decisions for individual students (Salvia & Ysseldyke, 1995). Salmon-Cox (as cited in Deno, 1985) found that teachers did not depend on norm-referenced achievement tests for making instructional decisions. The study also found that teachers relied more on their informal observations of students to make decisions about student performance. Salmon-Cox found a statistically significant discrepancy between actual student performance and teacher perception of student performance. These results suggested a need to avoid such discrepancies and to overcome the lack of support provided by the achievement tests for making ongoing instructional decisions. CBM evolved as a response to the limitations of standardized tests and has since been proved a reliable and valid measurement system to monitor progress.

The development of CBM measures was guided by the following underlying principles (Deno, 1985):

1. CBMs should be reliable and valid,
2. CBMs should be short and simple to administer,
3. Results should be easily understood, and
4. The measures should be inexpensive

Although early CBMs were developed for measuring progress in reading and writing, later years saw the development of CBMs in mathematics. Much needs to be done in the development of CBM in mathematics, especially for the secondary grades. Amidst the increasing need to monitor student progress at the secondary school level, Dr. Anne Foegen at Iowa State University started a project to develop CBMs for algebra. The study was called

Project AAIMS (Algebra Assessment and Instruction - Meeting Standards) and as part of the study, four algebra progress monitoring measures were developed. The four measures are Basic Skills, Algebra Foundations, Content Analysis, and Translations.

Project AAIMS was established to develop and validate a set of assessment tools that could be used in both general and special education settings to support increased student achievement in algebra for students with and without disabilities. Studies have been conducted to examine the reliability and criterion validity of the measures developed by Project AAIMS (e.g., Foegen, Olson, & Perkmen, 2005; Perkmen, Foegen, & Olson, 2006a, 2006b, 2006c).

Studies were also conducted to explore the extent to which these measures were sensitive to changes in student performance over time (Perkmen, Foegen, & Olson, 2006a, 2006b, 2006c). It was also investigated whether information gathered from the measures could be used to support teachers' instructional decision making and thereby enhance the learning of struggling students (Foegen & Olson, 2007). The Project AAIMS research program was designed to reflect the three stages that Fuchs (2004) asserted were necessary to "substantiate the tenability of measures for the purpose of progress monitoring" (p. 189). Fuchs' stages urge researchers and practitioners to investigate the technical adequacy of the measure at a single point of time (stage 1), to determine whether slopes indicate overall competence in the content area being assessed (stage 2), and to investigate whether the data obtained from the assessments can assist teachers' instructional decision making, thereby effecting gains in student achievement (stage 3).

In addition to examining the technical adequacy of the algebra measures, it is also important to examine whether the content of the measures corresponds to the specific curriculum that the schools are required to implement. Due to accountability requirements, it would also be desirable if the measures could predict student performance on high stakes tests. The present study addressed these concerns for two of the AAIMS measures.

Tests are tools that are often employed to assist in student evaluations (Matlock-Hetzel, 1997). As a basic unit of the test, the quality of each test item that constitutes the test plays an important role in deciding the nature and quality of the test. Item analysis serves to improve items to be used later in other tests, to eliminate ambiguous or misleading items in a single test administration, to increase instructors' skills in test construction, and to identify specific areas of course content which need greater emphasis or clarity (University of Washington, 2005). The quality of individual items is assessed by comparing students' item responses to their total test scores (University of Washington, 2005). The nature of the test items should be diagnostic in such a way that the test takers' performance on these items should indicate the extent of understanding, misunderstanding, or lack of understanding of the content of the test depending on the responses of the test takers. The most commonly used tools in test item analysis are item difficulty, item discrimination and differential item functioning. The present study investigated the item difficulty and item discrimination for items in two of the AAIMS measures.

This study contributes to the literature in the area of secondary mathematics progress monitoring by examining the extent to which the content tested in two of the Project AAIMS algebra measures, Basic Skills and Content Analysis, matches the content required by the Common Core State Standards. Using an existing data set, the study investigated the item difficulty and item discrimination statistics of the two progress monitoring measures. This study also explored the extent to which scores on subskill categories in each of the two progress monitoring measures predicted student performance on state achievement tests. Furthermore, this study investigated the predictive validity of scores obtained by grouping the subskill categories within the Common Core State Standards to students' scores on state achievement tests.

Research Questions

1. To what extent does the content tested in the two algebra progress monitoring measures align with the Common Core State Standards for Algebra?
2. What levels of item difficulty are represented in the skill/subskill categories of two algebra progress monitoring measures that correspond to the Common Core State Standards for Algebra? To what extent do the items discriminate the ability of the students?
3. To what extent do subtotal scores from the measures predict performance on state achievement tests in comparison to total scores?
    a. To what extent do subtotal scores based on the probe subskills (e.g., those used to develop the probes) predict state achievement test performance?
    b. To what extent to subtotal scores based on algebra standards (drawn from the alignment of the probes with the CCSS) predict state achievement performance?

Chapter 2: Literature Review

Described as a 'gateway' to higher mathematics, algebra is an important component in a student's learning, as competency in algebraic skills provides entry into many occupations and serves as a prerequisite for opportunities in many areas like postsecondary education (Stacey & Chick, 2004). Entry into many professional fields today requires knowledge of algebra. Employees must be able to use algebraic tools to translate problem situations involved in a given field to mathematical models that can be solved (Herscovics, 1989). In addition, algebra is used in nearly every scientific discipline. As algebra becomes increasingly important for employment, continued education, and daily living, all students must be successful in their ability to use algebra, not just students who are highly capable in mathematics.

Though algebraic thinking skills are introduced to students as early as prekindergarten, the National Council of Teachers of Mathematics (NCTM's) *Focal Points* document (NCTM, 2006) identifies algebra as an independent area of emphasis beginning in grade six. Traditionally, algebra is most often taught as an individual subject starting in grade 9, although some advanced students study algebra as early as grade 7 or 8. The growing emphasis on successful learning of algebra brings to the fore the need for and the importance of well researched and technically adequate assessment tools for measuring students' progress and proficiency in algebra. Project AAIMS (Algebra Assessment and Instruction - Meeting Standards; Foegen, 2003) has created algebra progress monitoring measures which meet these expectations.

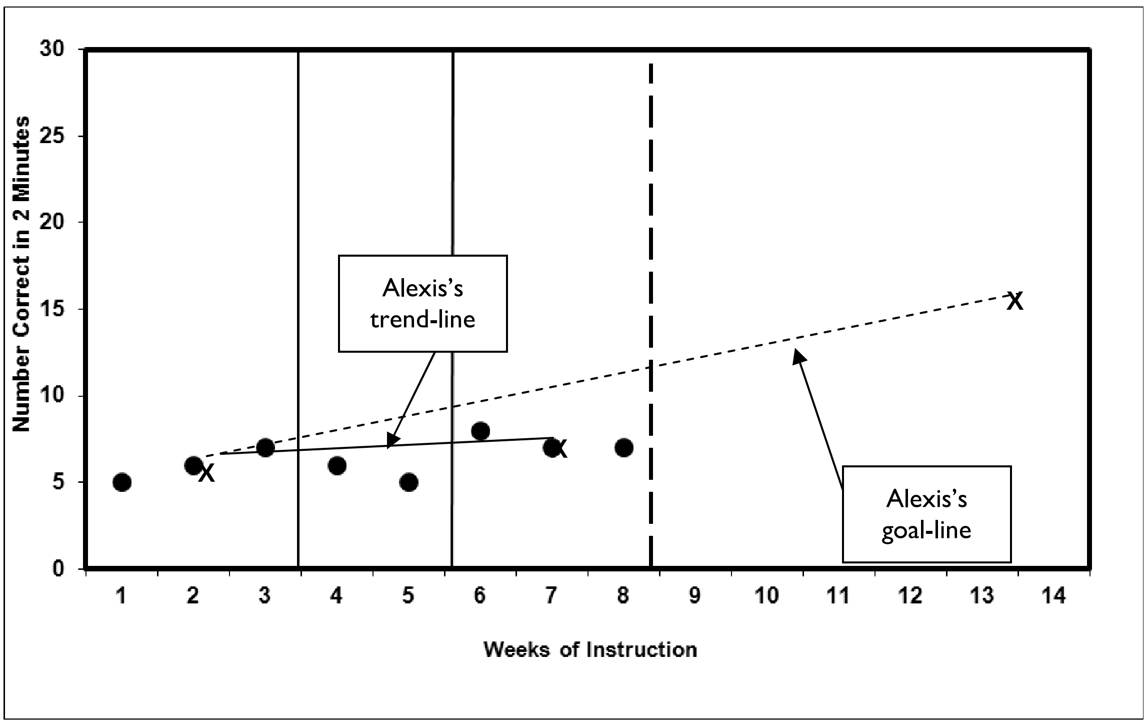
One of the goals of Project AAIMS was to develop and validate a set of assessment tools that could be used in both general and special education settings to support increased student achievement in algebra for students with and without disabilities. As part of the project, four algebra progress monitoring measures (APMMs) - Basic Skills, Algebra Foundations, Content Analysis, and Translations - were constructed and their technical adequacy was examined. These measures are described in detail in Chapter Three. Two of these APMMs, Basic Skills and Content Analysis, were investigated in this study to determine their content and predictive validity and their items' difficulty and discrimination levels.

The following sections provide information about the general concepts underlying Curriculum Based Measurement and a summary of the existing evidence for the APMMs. The chapter concludes with information about efforts to establish content validity in previous CBM research and procedures used to conduct item analysis.

Curriculum Based Measures

APMMs are a specific subset of curriculum based measures (CBMs). The functions and characteristics of APMMs or rather CBMs can be best explained using an example: Mr. B uses CBMs to monitor the progress of students in his class. He administers CBM probes once every week and graphs students' scores on individual student graphs. Figure 1 shows the CBM graph for a student named Alexis for an 8 week period. In the figure, the dotted line is Alexis' goal line. This goal line is drawn using her initial performance on the CBM probes and is the expected rate of growth. The bold line represents her current rate of performance and is called the trend line. The trend line is flatter than the goal line in the figure which indicates that Alexis is not performing according to the goal set for her and is unlikely to achieve her end of the year goal (or Alexis is not progressing at the expected rate). This prompts Mr. B to change the instructional plan. The vertical dotted line on the graph indicates the time when instructional plan was modified or changed. If, after the change in instructional plan, Alexis' trend line becomes steeper compared to her previous trend line and is approaching the goal line, it indicates that Alexis is improving and at some point she will reach the goal line. If the trend line continues to be flatter than the goal line, then Mr. B has to make a change again to improve Alexis' performance and start thinking about changing the rate of expected progress for Alexis. In this way, CBM helps teachers in manipulating their instructional plan or the rate of expected progress so as to achieve optimum results.

Figure 1. Alexis' CBM Computation Graph



Source:
http://www.studentprogress.org/summer_institute/2007/math/CBMMathHandouts_2007.doc

As was evident from this case study, CBM helps teachers to be proactive by constantly monitoring their students' performance. The visual display of students' performance in the form of graphs enables teachers to get a clear picture of their progress. Apart from that there are other features that make Curriculum Based Measurement a desirable form of assessment for educators; these were the guiding factors for the evolution of Curriculum Based Measurement (Deno, 2003).

- Short duration. The time taken to administer CBM probes is very short, ranging from 1 to 10 minutes. Administering these tests does not take much class time and thus does not affect the instruction time.
- Frequent administration. CBM probes are administered repeatedly and as frequently as once every week. This periodic and frequent administration facilitates teachers' early identification of students struggling to learn as well as informs teachers of the effectiveness of their instruction.
- Easy administration, scoring, and interpretation. Instructions for administering CBMs are well documented and simple to follow. Also, the administration and scoring procedures are standardized. This ensures good implementation fidelity and increases the likelihood that the results obtained from the tests are reliable. The simplicity of administration and scoring makes CBM easy for teachers and others involved in CBM implementation to use it properly and effectively. They can thus understand and use the obtained data for improved instruction and measured student performance.
- Multiple probes: To facilitate frequent administration, CBM uses multiple probes instead of identical probes to preclude students from memorizing the content. These probes are equivalent and have strong alternate form reliability.

- Technically adequate: One of the more salient features of CBM from researchers' points of view is technical adequacy. All the above features are more concerned with the practicality of administration. However, the results obtained will be useful only if the measures themselves are technically reliable and valid. It is also very important for the teachers to know that the tests they are using give reliable and valid results. One of the essential components while developing CBMs is that they should be technically adequate. CBMs are tested diligently for reliability, criterion validity, and sensitivity to growth before these probes are put to use.

All of these features in CBM enable teachers to figure out long term instructional goals and at the same time help them keep track of current achievement levels through easy to administer tests and simple to understand results. The following section provides information about the development and existing research evidence for the CBMs in algebra.

Algebra Progress Monitoring Measures

As described earlier in the chapter, Project AAIMS developed four algebra progress monitoring measures. They are Algebra Basic Skills (ABS), Algebra Foundations (AF), Algebra Content Analysis (ACA), and Translations. Sixteen technical reports were released that investigated these measures for technical adequacy (see http://www.ci.hs.iastate.edu/aaims/technical_reports.php). Currently, ABS and ACA are most frequently selected by teachers for use in their classrooms. As part of Project AAIMS, studies were conducted to establish and replicate evidence of the reliability and validity for APMMs, in particular ABS and ACA. In the following sections, I summarize the findings from the latest five technical reports for Project AAIMS (Foegen, Olson, & Perkmen, 2005, Perkmen, Foegen, Olson, 2006a, 2006b, 2006c; Foegen & Olson, 2007)

Alternate form reliability for Algebra Basic Skills (ABS), documented across three studies, ranged from .49 to .91. Correlations mostly ranged from .80 to .90 except in one study (Foegen & Olson, 2007) where a low correlation range was attributed to the limited number of class types for ABS (primarily remedial courses) in that study restricting the range of obtained scores, and lowering the correlations. Test retest reliability for ABS ranged from .75 - .89. For Algebra Content Analysis, across four studies, alternate form reliability ranged from .48 to .94. Lower correlations in this case were observed in the scores obtained in the beginning of the school year. Correlations were greater than .80 for the second half of the school year. Test retest reliability for ACA ranged from .64 - .88. In all, correlation values indicated that ABS and ACA are sufficiently reliable measures.

Concurrent validity correlations between the ABS and the Iowa Tests of Educational Development (ITED) Computation subtest were not significant across two studies (Foegen & Olson, 2007). For ACA, concurrent validity correlation between the ITED Computation subtest, the ITED Concepts/Problem Solving subtest , and the Iowa Tests of Basic Skills (ITBS) Math Total scale were .79, .62, and .30, respectively, for one study (Perkmen, Foegen, & Olson, 2006). Another study (Foegen & Olson, 2007) produced values of .39 and .36 with ITED Computation and ITED Concepts/Problem Solving respectively. The correlation with ITBS Math Total was not significant; this result may have been associated with the small sample size (N = 21) of eighth grade students in the sample who were taking algebra for high school credit.

The studies also examined the validity of the APMMs administered at the beginning of a course for predicting students' later performance on state achievement tests. Predictive validity for ABS ranged from .33 to .40 with ITED Computation and from .36 to .45 for

ITED Concepts/Problem Solving. For ACA, predictive validity ranged from .32 to .42 with ITED Computation and from .25 to .30 with ITED Concepts/Problem Solving.

The absence of strong concurrent and predictive validity in ABS and ACA with ITED and ITBS might be because of the differences in target content for the two types of assessments. Whereas the APMMs were specifically developed for algebra, the objective for ITED is to test students' proficiency in generally held mathematics objectives in high school. Thus, limited numbers of items in the tests belong to any one specific course. As a result, the scores on ITED represent an overall mathematics competency in students rather than proficiency in any particular course.

CBMs have been distinctive from other forms of classroom assessment in that their development includes research on the measures' technical adequacy. Another important feature of CBMs is found in the name of the measures; that is, they are "curriculum based." The following section outlines the basic concept of content validity and summarizes the evidence of content validity for mathematics CBMs.

Content Validity

Content validity is one of the three essential validity measures used in test construction (the other two being criterion validity and construct validity). Content validity is a measure of the extent to which a test covers the content it is testing (Carmines & Zeller, 1991). Content validity of a test is usually reported in the 'development' section of the test's manual which includes the processes for item development and selection (Salvia & Ysseldyke, 2007). This section also describes the sources used to develop test items. For example, items in ITBS and ITED were developed using curriculum guides and textbooks. Teachers and administrators were consulted in the writing of test items (Salvia & Ysseldyke, 2007). The Group Mathematics Assessment and Diagnostic Evaluation (G.MADE) is another norm referenced, group administered test for assessing mathematics skills for grades K-12. The content for this test is based on the NCTM standards and the items were developed based on state standards, curriculum benchmarks, math textbooks and research on best practices in mathematics teaching (Salvia & Ysseldyke, 2007). Though details such as those given above are included in the tests' manuals as part of establishing their content validity, there are usually no formal studies or proven processes that establish the adequacy of content validity. As such, the word of the author is taken as the criteria for efficacy for content validity.

Content validity in mathematics CBM. One of the criteria for developing CBMs is that they are created using local curriculum and are very much connected to the curriculum of instruction. Deno (1985) noted that at the time CBM was developed, the content of many standardized achievement tests represented generally held expectations for proficiency in a content area, but did not reflect local instructional content effectively. By drawing from the local instructional curriculum and materials for tasks and content, CBM data provided teachers with greater confidence that students' scores were representative of proficiency in the local curriculum. As a result, the literature on CBM includes little formal attention to gathering evidence of the measures' content validity.

A review of the technical adequacy literature for CBMs in mathematics identifies two examples of efforts to directly attend to the content validity of the measures. At the elementary level, the work of Lynn and Doug Fuchs of Vanderbilt University included the development of two types of mathematics CBMs based on state curriculum guidelines. The measures, which address computation and concepts/applications, used the Tennessee mathematics curriculum at the time of test development and teachers' feedback for developing the probes (Fuchs, Hamlett, & Fuchs, 1998). The curriculum was analyzed to determine the most critical skills and concepts at each grade level, items representing these skills and concepts were developed, and teachers provided feedback on the appropriateness of the items for representing the instructional curriculum.

At the secondary level, the development of ACA involved examining the content from a conventional algebra textbook that had been adopted by all the districts participating in the development of the APMMs (Foegen, Olson, & Impecoven-Lind, 2008). Similar to the process used by Fuchs et al. (1998), the content of the textbook was evaluated to identify a small number of critical concepts in each chapter and items were developed to reflect these skills and concepts. Teachers participating in the project reviewed the listing of critical skills and concepts, along with the items, and provided feedback used to revise and refine the items and the measures. Feedback was also gathered from faculty in mathematics education before the items were finalized.

With the Common Core State Standards (CCSS; Common Core State Standards Initiative. n.d.) being adopted by forty one states, schools will soon be teaching content recommended by these state standards. In this context, it becomes important to know whether the assessments being used in schools are testing the content being taught. Because the APMMs are currently being used in many schools it is imperative to investigate whether the content in these measures aligns with the content in CCSS.

The features of test development discussed thus far (reliability, criterion validity, and content validity) represent traditional constructs associated with classical test theory (Kline, 1986). These constructs place primary emphasis on the total score derived from the assessment. Another tool often used in the development of achievement tests is item analysis, which examines the quality and contributions of individual items. The following section provides more information about item analysis.

Item Analysis

The quality of an item decides the quality of the test. Classical item analysis helps in improving the quality of tests by revising and improving the items in the test (Livingston, 2006). Item difficulty is one of the statistics in classical item analysis. In a test it is important to know whether the difficulty of an item is suited to the level of students for whom the test is intended.

Item difficulty is the proportion of students taking the test who attempted that item successfully. The higher the value, the easier the item is. Item difficulty ranges from 0 to 1. In traditional achievement tests, items displaying values closer to 0 (indicating that almost all students got the item wrong) and 1 (indicating that almost everyone got the item correct) should be revised or removed, because they offer little ability to discriminate among students at varying proficiency levels. Items having difficulty ranges from .2 to .8 provide the maximum information about proficiency among students. There is an exception to this theory when the tests are used to assess students of an extreme group. For example, in a special education scenario, the teacher would be looking for tests that have easy items because in such a case, students are unlikely to attempt difficult items successfully and so items of higher difficulty ranges would not provide much information about student ability.

Item discrimination is the other statistic in classical item analysis. The item discrimination index indicates whether items are discriminating students based on their ability to perform (Allen & Yen, 1979). That is, the item is able to distinguish between high and low performing students. Item discrimination ranges from 0 to 100%. If all those in the upper group answered correctly and all those in the lower group answered incorrectly, then the discrimination index would be 100%. Zero discrimination occurs when equal numbers in both groups answer correctly. Negative discrimination occurs when more students in the lower group answer correctly than the upper group. Allen and Yen suggested a scale for interpreting item discrimination in which items with negative values are judged unacceptable (and should be checked for errors) and those with discrimination values between 0% and 24% are potential candidates for approval. Items with discrimination values from 25% to 39% are considered good items, and those with values at or above 40% are judged to be excellent items (Findley,1956).

A review of the CBM literature on technical adequacy did not produce any study that did classical item analysis with CBM. CBMs measure growth and as such they are designed to avoid ceiling scores (so that the tests continue to show growth), unlike the traditional tests where ceiling scores would be desired (indicating successful instruction). As a result of the intent to avoid ceilings, many items in CBMs remain unattempted, which makes the task of item analysis difficult. This study explored the processes of doing item analysis on two algebra progress monitoring measures to examine item level difficulty and discrimination.

Predictive Validity

With regard to predictive validity, there are a few studies in the literature related to CBM. Again the field is very narrow with regard to CBMs in mathematics. In a study by Singamaneni, Foegen, and Olson (2009), it was established that the Early Numeracy Indicators (math CBMs) for grades K-1 were able to predict student performance on third grade ITBS from kindergarten and first grade ENI performance. In another study, Shapiro, Keller, Lutz, Santoro and Hintze (2006) found that CBM measures of reading, Math

Computation, and Math Concepts and Applications had moderate to strong correlations with state assessment tests.

With the scenario in Iowa schools changing from complete local autonomy in regards to curriculum selection, to adopting Common Core State Standards (CCSS), it is important to see whether the content of the tests teachers are using align with CCSS. APMMs were constructed in accordance with the locally used traditional textbooks. In keeping with the current changes in the state's policy to adopt CCSS, it is important to establish the content validity of APMM with regard to CCSS. This study explores the content validity of two APMMs with CCSS.

Reliablity and validity for APMM were established in earlier technical adequacy reports as described in the above sections. These pertain to the quality of the measures as a whole. But analysis at the item level is yet to be taken up. This study explores classical item analysis for two of the APPMs. Item difficulty and item discrimination for ABS and ACA were investigated in this study. After exploring the quality of items, this study looked into the predictive power of these two measures in predicting student performance in ITBS and ITED tests. Though the predictive validity of the total scores from these two measures have been investigated and reported in earlier technical reports as described in the sections above, I was interested in determining whether the subskill categories in each of these two measures can predict performance in ITBS and ITED. The present study is based on the three research questions listed below.

Research Questions

1. To what extent does the content tested in the two algebra progress monitoring measures align with the Common Core State Standards for Algebra?

2. What levels of item difficulty are represented in the skill/subskill categories of two algebra progress monitoring measures that correspond to the Common Core State Standards for Algebra? To what extent do the items discriminate the ability of the students?
3. To what extent do subtotal scores from the measures predict performance on state achievement tests in comparison to total scores?
    a. To what extent do subtotal scores based on the probe subskills (e.g., those used to develop the probes) predict state achievement test performance?
    b. To what extent to subtotal scores based on algebra standards (drawn from the alignment of the probes with the CCSS) predict state achievement performance?

Chapter 3: Method

This chapter is divided into four sections. The first section describes the participants and settings. The second section describes the measures used in this study. The third section describes the procedures used in the original study to generate the extant data used for the current study. The final section describes the procedures used to investigate the research questions.

The data used for this study were originally collected as part of Project AAIMS during the academic year 2006-2007 (Foegen & Olson, 2007). For the original study (Foegen & Olson, 2007a), written parental/guardian consent and written student assent were obtained for all the student participants in accordance with Iowa State University's Human Subjects Review Committee. IRB approval has been obtained to do further analysis of these data for the purpose of this study (See Appendix A).

Participants and Settings

The data for this study were taken from the data collected for an AAIMS study during the academic year 2006 -2007. Participants for this study were students from two districts, identified as District A and District B.

Eighty three students from District A and fifty one students from District B participated in this study. The data were collected by two teachers in District A and three teachers in District B. Demographic data by district for participating students are presented in Table 1.

Table 1
Demographic characteristics of student participants

	Gender		Ethnicity				Free/reduced lunch	Sped	N
	Female	Male	Black	White	Asian	Hispanic			
District A	45	38	1	81	1	0	15	8	83
District B	29	22	4	44	0	1	Not reported by the district	9	51

Note Sped = Special Education

Students participating in the study were enrolled in one of four types of algebra classes. A total of 67 students were participating in a traditional Algebra 1 course taught using a conventional time frame (one year for District A with 45 minute periods, and one half year for District B, using block scheduling with daily 90 minute periods). Of these, 22 were 8th grade students in District A completing a high school algebra course; these students, who comprised a single class, were identified as advanced in mathematics within their district. The remaining 45 students were enrolled in one of four different sections of Algebra 1. All of the Algebra 1 students were from District A. The remaining 16 students from District A and all the students from District B were enrolled in one of six sections of Algebra 1A. Algebra 1A is a course in which the first half of a traditional Algebra 1 course is taught in the conventional time frame.

Measures

For this study, two of the four AAIMS measures were investigated: Algebra Basic Skills and Algebra Content Analysis. The other two measures developed and studied in Project AAIMS were the Algebra Foundations and the Translations measures, but data from these measures were not considered for the present study. In addition, the criterion measures used in the original study included the Iowa Test of Basic Skills (ITBS) and the Iowa Tests of Educational Development (ITED). The following sections describe the measures from the original study.

Algebra Basic Skills (ABS). This measure assesses the skills that students are expected to have acquired for automaticity in algebra. The ABS measure addresses proficiency in the skills of solving simple equations, applying the distributive property, working with integers, combining like terms and applying proportional reasoning. This probe has 60 constructed response items and students have five minutes to work on it. Each item that is answered correctly gets a score of one point. A copy of an Algebra Basic Skills measure is presented in Appendix B. Technical adequacy for ABS was documented in the technical reports as part of Project AAIMS. The alternate form reliability estimates ranged from .81 - .91 (Perkmen, Foegen, & Olson, 2006a, 2006b) and .49 - .90 (Foegen & Olson, 2007). The lower results for the later study were attributed to the lower range of scores due to limited class types participating in the study (Foegen & Olson, 2007). The test retest reliability estimates ranged from .75-.89 (Perkmen et al, 2006a; 2006b). Predictive validity estimates for ABS ranged from .33 - .40 with ITED Computation and from .36 - .45 with ITED Concept/ Problem Solving (Perkmen et al., 2006a; 2006b).

Algebra Content Analysis (ACA). This measure assesses key concepts from the first two-thirds of a traditional algebra course. This probe has 16 multiple choice items and students get 7 minutes to work on it. In addition to choosing the right answer, students are encouraged to show their work in order to earn partial credit in the event that they do not select the correct answer. Scoring for the ACA probes is done by comparing student responses to a rubric-based key created by the research staff. Each of the 16 problems is worth up to three points. Students earn full credit (three points) by circling the correct answer from among the four alternatives. If students circle an incorrect response and do not show any work, their answer is considered a ‘guess;’ the total number of guesses is recorded for each probe and subtracted from the points earned on the other items. In cases where students show work, the scorer compares the student’s work to the rubric-based key, and determines whether the student has earned 0, 1, or 2 points of partial credit. The number of points earned across all 16 problems and the number of guesses are recorded. A final score is computed by subtracting the number of guesses from the total number of points earned on the probe. A copy of an Algebra Content Analysis measure is presented in Appendix C.

Technical adequacy for ACA was established as part of Project AAIMS and data were reported in the AAIMS technical reports. Alternate form reliability estimates ranged from .48 - .94 (Foegen & Olson, 2007; Perkmen et al, 2006a; 2006b; 2006c). The authors observed that the lower estimates were obtained from the scores collected in the first administrations that were at the beginning of the school year. Test retest reliability estimates ranged from .64 - .88 (Perkmen et al, 2006a; 2006b; 2006c). Concurrent validity estimates for ACA were .79 with ITED Computation, .62 with ITED Concepts/ Problem Solving, and .30 with ITBS Math Total (Perkmen et al, 2006c). Predictive validity estimates for ACA ranged from .32-.42 with ITED Computation, and from .25 - .30 with ITED Concepts/ Problem Solving (Perkmen et al., 2006a; 2006b).

Criterion Measures. The criterion measures used for this study were the Iowa Tests of Basic Skills (ITBS) and the Iowa Tests of Educational Development (ITED). ITBS is a norm referenced, group administered battery of tests for grades K to 8. This test is used in Iowa for accountability for Annual Yearly Progress and provides a comprehensive assessment of student progress in major content areas (Hoover et al., 2001). The total score in each content area is the average of the scores on all the subtests for that content and the scores are reported in standard scores and percentile ranks. There are three subtests for testing mathematics content namely, Mathematics Concepts and Estimation, Mathematics Problem Solving and Data Interpretation, and an optional Mathematics Computation subtest. The Mathematics Total scale score provides a composite estimate of student proficiency in mathematics.

As with ITBS, ITED is also a norm referenced group administered battery of tests for grades 9 through 12 and the results are used for accountability in schools' Annual Yearly Progress report. ITED has tests for English language, mathematics, and science. For mathematics content there are two subtests namely, (a) Concepts and Problem Solving and (b) Computation. While the Concepts and Problem Solving subtest measures “students' abilities to use appropriate mathematical reasoning” (Iowa Testing Program, n. d.), the Computation subtest measures skills “related to the computational manipulations needed throughout the secondary school mathematics curriculum” (Iowa Testing Program, n. d). The Concepts and Problem Solving score is also reported by the test developers as the Mathematics Total score for the measure.

Procedures for the Original Study

As part of the original study (Foegen & Olson, 2007a, 2007b), Project AAIMS research staff visited each class at the beginning of the school year (District A) or semester (District B) to present information about the project and gather informed consent. During the period of study, four probes were administered each month. Administration of the probes was not identical across teachers, districts, or measures. Details about the types of measures administered by each participating teacher are provided in Table 2. Though teachers were given the option to choose any of the three measures to monitor their students' progress, the most frequently selected measure was Algebra Content Analysis, followed by the Algebra Basic Skills measure. None of the teachers chose to administer the Algebra Foundations measure.

Procedures for the Current Study

Procedure for Research Question 1. The first research question deals with investigating the alignment of the content tested in the ABS and ACA measures with the content of the Common Core State Standards (CCSS) for Algebra (Common Core State Standard Initiative, n.d.). To accomplish this task, the categorization of skills from the measure development templates from both ABS and ACA were aligned with the skill categories of the four CCSS for high school algebra.

Table 2

Details on measures administered by teacher

District	Teacher	Number of participants	Period/Block	Probe
A	1	64	2	Algebra Content Analysis
			3	Algebra Content Analysis
			4	Algebra Content Analysis
			6	Algebra Content Analysis
			7	Algebra Content Analysis
A	2	19	5	Basic Skills, Algebra Content Analysis
			7	Basic Skills, Algebra Content Analysis
B	3	22	1	Algebra Content Analysis
			2	Algebra Content Analysis
B	4	18	2	Algebra Content Analysis
B	5	11	2	Basic Skills, Algebra Content Analysis

Table 3 shows the CCSS for high school algebra where phrases in the first column indicate Domains, or larger groups of related standards. The phrases across from each Domain in the adjacent column are standards that define what students should understand and be able to do.

Table 3

Common Core State Standards for high school algebra

Standard Domains	Standards in Detail
CCSS 1. Seeing Structure in Expressions	• Interpret the structure of expressions • Write expressions in equivalent forms to solve problems
CCSS 2. Arithmetic with Polynomials and Rational Expressions	• Perform arithmetic operations on polynomials • Understand the relationship between zeros and factors of polynomials • Use polynomial identities to solve problems • Rewrite rational expressions
CCSS 3.Creating Equations	• Create equations that describe numbers or relationships
CCSS 4. Reasoning with Equations and Inequalities	• Understand solving equations as a process of reasoning and explain the reasoning • Solve equations and inequalities in one variable • Solve systems of equations • Represent and solve equations and inequalities graphically

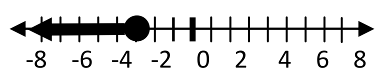
Table 4 shows skills and subskills in ACA. To address Research Question 1, I aligned the ACA subskills to the CCSS high school algebra domains and their respective standards. Though there is a more detailed explanation for each standard in CCSS, I decided to use the standards' domains for alignment. I chose this organization because the CCSS for high school algebra brings together all the standards covered for grades 9 – 12 (e.g., both Algebra 1 and Algebra 2), whereas the AAIMS measures were designed to test skills acquired in

Algebra 1 and also include some pre-algebra skills. Hence, the content covered in CCSS will be far more advanced than that in ACA. Also, because the CCSS do not specify standards by courses (like Pre Algebra or Algebra 1), it is not possible to isolate standards for Algebra 1. As a result, the subskills in ACA will not align perfectly with the CCSS for high school algebra.

Table 4
Algebra Content Analysis skills and subskills

Skills Tested	Subskills
ACA 1 Connections to Algebra	ACA 1 Evaluate expressions that include exponents and order of operations with given values Sample problem - Evaluate $a^2 - b \div 2$ when a = 4 and b = 6
ACA 2 Properties of Real Numbers	ACA 2.1 Simplify expressions that include integers and combination of like terms Sample problem - Simplify: 9r + 3r – 3 + r2 + 2 ACA 2.2 Simplify expressions that include integers and combination of like terms and application of the distributive property (1 addition, 1 subtraction) Sample problem - Simplify: 4(n – 2) + 2(n + 6)
ACA 3 Solving Linear Equations	ACA 3.1 Solve linear equations with 2 steps Sample problem - Solve: 3x – 4 = 20 ACA 3.2 Solve equations with variables on both sides Sample problem - Solve: 5z + 4 = – 3z – 12
ACA 4 Graphing Linear Equations & Functions	ACA 4.1 Identify a line on a graph Sample problem - Which line on the graph is y = 2? ACA 4.2 Find the slope of a line through 2 points Sample problem - Find the slope of a line through (1, 3), (2, 5)

Table 4 (Continued)

Skills Tested	Subskills
ACA 5 Writing Linear Equations	ACA 5.1 Slope-intercept form Sample problem - Write the equation in slope-intercept form: $m = \frac{1}{2}$ $b = 3$ ACA 5.2 Write equation for line through 2 points Alternate between point-slope (ACA 5.2a) and slope-intercept form (ACA 5.2b) ACA 5.2a Sample problem - Write the equation of a line through (5, 3) (4, 9). Use point-slope form. ACA 5.2b Sample problem - Write the equation of a line through (4, 2) (6, 3). Use slope-intercept form.
ACA 6 Solving & Graphing Linear Inequalities	ACA 6 Interpret a graph of an inequality Sample problem - This graph shows the solution for which equation? 
ACA 7 Systems of Linear Equations & Inequalities	ACA 7.1 Solve linear system by substitution Sample problem - Solve the linear system: x – y = 4 x + 2y = 19 ACA 7.2 Solve linear system by linear combination Sample problem - Solve the linear system: – 6x + 3y = – 6 2x + 6y = 30
ACA 8 Exponents & Exponential Functions	ACA 8.1 Evaluate expressions with exponents (either a negative base or a negative exponent - second or third power) Sample problem - Evaluate the expression: 4^{-2} ACA 8.2 Simplify expressions with exponents Sample problem - Simplify the expression: $\frac{a^2}{ab^3} \bullet \frac{b^4}{a^3}$

In this scenario, I have applied a decision rule that if a subskill in ACA relates to a broader standard in CCSS for algebra, then that subskill is testing at least a part of that particular standard in CCSS (indicating that the basic skills required to cover the more advanced levels of the CCSS are covered by ACA). For example, in ACA 1 the skill tested is Connections to Algebra and the subskill tested is to evaluate expressions that include exponents and order of operations with given values. Two CCSS standards align with this skill. They are Seeing Structure in Expressions (where students interpret the structure of expressions and write expressions in equivalent forms to solve problems) and Arithmetic with Polynomials and Rational Expressions (where students perform arithmetic operations on polynomials).

Table 5 shows the skills and subskills in ABS. These skills are more representative of pre- algebra skills than those learned in a typical Algebra 1 course as indicated by the name of the measure, that is, Algebra Basic Skills. As a result there are some test items in ABS that test arithmetic skills and hence do not align with any of the algebra CCSS. Nevertheless, these skills are essential for students to acquire competency in algebraic skills, as command of arithmetic skills forms the basic foundation for acquiring any higher order mathematics skills. I have used the same method and applied the same decision rule for ABS as with ACA for alignment with CCSS for high school algebra. For example, ABS 1 tests the skill Solving Simple Equations (Basic facts) and the CCSS that aligns with this skill is Reasoning with Equations and Inequalities, which expects students to understand solving equations as a process of reasoning and explain the reasoning and to solve equations and inequalities in one variable.

Table 5

Algebra Basic Skills skills and subskills

Skills Tested	Subskills and Sample Problems
ABS 1 Solving simple equations (Basic facts)	ABS 1.1 Solve simple addition equations Sample problem - 6 + p = 11
	ABS 1.2 Solve simple subtraction equations Sample problem - 12 – e = 4
	ABS 1.3 Solve simple multiplication equations Sample problem - 4r = 28
	ABS 1.4 Solve simple division equations Sample problem - 63 ÷ c = 9
ABS 2 Applying the distributive property	ABS 2.1 Apply the distributive property and add or subtract an integer Sample problem - 4(3 + s) – 7
	ABS 2.2 Apply the distributive property and add or subtract a variable Sample problem - 5(b – 3) – b
	ABS 2.3 Apply the distributive property and add or subtract an integer and a variable Sample problem - 5(3 + f) – 2f + 6
	ABS 2.4 Apply the distributive property to get a quadratic Sample problem - 2 + w(w – 5)
ABS 3 Working with integers	Sample problems - – 5 + 6 – 6 OR 9 + (– 3) – 8 (all only have 3 integers)
ABS 4 Combining like terms	ABS 4.1 Add and/or subtract linear terms with the same variable Sample problem - b + b + 2b
	ABS 4.2 Add and subtract linear variables and integers Sample problem - 3z – 8z + 2 + 9

Table 5 (Continued)

Skills Tested	Subskills and Sample Problems
	ABS 4.3 Add and subtract quadratics and integers Sample problem - – 3w2 + 5w2 – 5 + 12
	ABS 4. 4 Add and subtract linear and quadratic terms Sample problem - y2 + y – 4y + 3y2
	ABS 4.5 Add and subtract linear terms, quadratic terms, and integers Sample problem - 6a + 2a – 9 + 3a2
ABS 5 Applying Proportional Reasoning	ABS 5.1 Calculate equivalent fractions Sample problem - $\frac{a^2}{ab^3} \bullet \frac{b^4}{a^3}$
	ABS 5.2 Make measurement conversions Sample problem - 4 quarts = 1 gallon ____ quarts = 3 ¼ gallons

Procedures for Research Questions 2 and 3. For Research Questions 2 and 3, the data used consisted of two equivalent probes of ABS and eight equivalent probes of ACA that were administered to the students. Alternate form reliabilities for these multiple probes were established as part of examining the technical adequacy of both ABS and ACA (Perkmen et al., 2006a, 2006b). A total of 35 probes for ABS that were administered on September 13 and November 3 of 2006 were used for this analysis. Similarly 134 probes for ACA that were administered from October 23 to December 14 of 2006 were used for this analysis.

Research Question 2 deals with the item difficulty levels in the skill/subskill categories of ABS and ACA and the item discrimination of both these measures. To find the item difficulty level, I calculated the percentage of students who answered the items in each of the skill/subskill categories correctly**.**

Also I wanted to investigate the difficulty levels within the four CCSS domains for high school algebra. For this purpose, I aligned all the skill/subskill categories of ABS and ACA with the four CCSS domains and conducted the analyses within these subskill groupings. Tables 6 and 7 display the groupings of the ACA and ABS subskills, respectively, by the CCSS domains. Some of the skill/subskill categories came under more than one CCSS domain for high school algebra and some CCSS domains did not have alignment with any of the skill/subskill categories of the progress monitoring measures. For example in ACA (Table 6), Skill 1 (Evaluate expressions that include exponents and order of operations with given values) comes under both CCSS domain 1 and domain 2, that is 'Seeing Structure in Expressions' (domain 1) and 'Arithmetic with Polynomials and Rational Expressions' (domain 2)**.** Also**,** this grouping resulted in the exclusion of some skill categories. For example, ACA Skill 8 (exponents and exponential functions) does not come under any of the four CCSS domains. I assert that a test should be assessing all the aspects of the criterion under consideration (in this case the four CCSS domains), but at the same time it can also test more than the criterion.

In ABS (Table 7) however, domain 3, that is, 'Creating Equations' does not match/align with any ABS skill categories. As mentioned earlier, the purpose of the ABS measure was to test basic algebra skills and so it does not include many higher level skills like creating equations that students are required to acquire in their algebra courses. In

Chapter 4, the content of Tables 6 and 7 will be used to organize the results of the item analyses for Research Question 2.

Table 6
Skill categories in ACA grouped by CCSS for high school algebra

CCSS	Skill categories in ACA
1. Seeing Structure in Expressions	ACA 1 Evaluate expressions that include exponents and order of operations with given values
	ACA 2.1 Simplify expressions that include integers and combination of like terms
	ACA 2.2 Simplify expressions that include integers and combination of like terms and application of the distributive property (1 addition, 1 subtraction)
	ACA 8.2 Simplify expressions with exponents
2. Arithmetic with Polynomials and Rational Expressions	ACA 1 Evaluate expressions that include exponents and order of operations with given values
	ACA 2.1 Simplify expressions that include integers and combination of like terms
	ACA 2.2 Simplify expressions that include integers and combination of like terms and application of the distributive property (1 addition, 1 subtraction)
3. Creating Equations	ACA 4.1 Identify a line on a graph
	ACA 4.2 Find the slope of a line through 2 points
	ACA 5.1 Slope-intercept form
	ACA 5.2 Write equation for line through 2 points

Table 6 (Continued)

CCSS	Skill categories in ACA
4. Reasoning with Equations and Inequalities	ACA 3.1 Solve linear equations with 2 steps
	ACA 3.2 Solve equations with variables on both sides
	ACA 4.1 Identify a line on a graph
	ACA 4.2 Find the slope of a line through 2 points
	ACA 5.1 Slope-intercept form
	ACA 5.2 Write equation for line through 2 points
	ACA 6 Interpret a graph of an inequality
	ACA 7.1 Solve linear system by substitution
	ACA 7.2 Solve linear system by linear combination

I used LERTAP's proportional method to find item difficulty. LERTAP is the acronym for the Laboratory of Educational Research Test Analysis Package. It is a system for item, test, and survey analysis. First developed in 1971-72, today it is in its fifth version and is used in Canada and the United States. Though the method to calculate item difficulty in LERTAP (Item difficulty calculations, n.d.) remains the same (percent of students who successfully attempted the item), it is the way item scores are handled that makes it more suitable for this study. This method counts any response as being correct if its corresponding weight is greater than zero. This method does not take into account any differences that may exist among response weights. In ACA, partial credit is given for items that are partly correct and negative scoring is done for incorrect guessing (since this measure has multiple choice answers) if the student has not shown any work. This scoring method was adopted to discourage students from guessing or randomly selecting the answers. Using LERTAP's method, I recoded all scores greater than zero as correct and those equal to or less than zero as incorrect. Because of the large number of items in ABS (60), I grouped the ABS items by subskills and coded the subskill categories as correct if the student attempted even one item in that category successfully.

Table 7

Skill categories in ABS grouped by CCSS for high school algebra

CCSS	Skill categories in ACA
1. Seeing Structure in Expressions	ABS 2 Applying the distributive property
	ABS 4 Combining like terms
2. Arithmetic with Polynomials and Rational Expressions	ABS 2 Applying the distributive property
	ABS 4 Combining like terms
3. Creating Equations	None of the skill categories fit here
4. Reasoning with Equations and Inequalities	ABS 1 Solving simple equations (Basic facts)

Item discrimination must be computed at the item level. I rank ordered all the students by their total scores and grouped the top and bottom thirty percent of students. I calculated the percent of students who got the item correct in both groups and used the following formula:

Item Discrimination = (Upper Group % Correct) – (Lower Group % Correct).

Tables 6 and 7 will also be used in Chapter 4 to report the results of these analyses using the CCSS Domains as groups.

Research Question 3 investigated the predictive validity of the algebra measures with respect to student performance on ITBS/ITED scores. More specifically, I examined whether correlations between ABS and ACA scores based on the CCSS skills categories were more or less predictive than scores for the total ABS and ACA measures. Correlations were also computed between combined scores for items within each skills category and ITBS/ITED scores to find the predictive validity for each of the subskill categories. Also I examined the predictive power of combined scores when items are grouped by CCSS Domains. For all of these analyses, Pearson product moment correlation coefficients were computed.

Chapter 4: Results

This chapter discusses the results by research questions. Within each research question, the results are detailed by probe type.

Research Question 1

Because the Common Core State Standards (CCSS) have been adopted by Iowa and many other states, it is important to see whether the tests that are being used in schools will continue to maintain their content validity. Within this context, the content tested in the two algebra progress monitoring measures (APMMs) was examined to determine alignment with the Common Core State Standards (CCSS) for algebra. Results of this investigation are described below. Results for Algebra Content Analysis (ACA) will be discussed first followed by those for Algebra Basic Skills (ABS)

Algebra Content Analysis. As described in the method chapter, I aligned the ACA subskills to the CCSS high school algebra domains and their respective standards. Table 8 shows the skills and subskills in ACA and how they correspond to the Common Core State Standards (CCSS) for high school algebra. In the column for CCSS high school algebra, domains and the respective standards that best align with the subskills of ACA are arranged. As can be seen from Table 8, all the skill categories in ACA align with at least one or more CCSS high school algebra domains.

Algebra Basic Skills. I used the same process of alignment for ABS that I used for ACA. Table 9 shows the alignment of ABS skill categories with those of CCSS-high school algebra domains and their respective standards. As can be seen in Table 9, skill categories ABS 3 and ABS 5 do not align with any of the CCSS high school algebra domains, while skill categories ABS 1, ABS 2, and ABS 4 correspond with at least one of the CCSS high school algebra domains.

Table 8
Algebra Content Analysis skills and subskills alignment with the high school algebra Common Core State Standards

Algebra Content Analysis		Common Core State Standards - High School Algebra
Skills Tested	Subskills	
ACA 1 Connections to Algebra	ACA 1 Evaluate expressions that include exponents and order of operations with given values Evaluate $a^2 - b \div 2$ when a = 4 and b = 6	**Seeing Structure in Expressions** • Interpret the structure of expressions • Write expressions in equivalent forms to solve problems **Arithmetic with Polynomials and Rational Expressions** • Perform arithmetic operations on polynomials
ACA 2 Properties of Real Numbers	ACA 2.1 Simplify expressions that include integers and combination of like terms Simplify: 9r + 3r – 3 + r2 + 2	**Seeing Structure in Expressions** • Interpret the structure of expressions • Write expressions in equivalent forms to solve problems **Arithmetic with Polynomials and Rational Expressions** • Perform arithmetic operations on polynomials

Table 8 (Continued)

Algebra Content Analysis		Common Core State Standards - High School Algebra
	ACA 2.2 Simplify expressions that include integers and combination of like terms and application of the distributive property (1 addition, 1 subtraction) Simplify: 4(n – 2) + 2(n + 6)	**Seeing Structure in Expressions** • Interpret the structure of expressions • Write expressions in equivalent forms to solve problems **Arithmetic with Polynomials and Rational Expressions** • Perform arithmetic operations on polynomials
ACA 3 Solving Linear Equations	ACA 3.1 Solve linear equations with 2 steps Solve: 3x – 4 = 20	**Reasoning with Equations and Inequalities** • Understand solving equations as a process of reasoning and explain the reasoning • Solve equations and inequalities in one variable
	ACA 3.2 Solve equations with variables on both sides Solve: 5z + 4 = – 3z – 12	**Reasoning with Equations and Inequalities** • Understand solving equations as a process of reasoning and explain the reasoning • Solve equations and inequalities in one variable

Table 8 (Continued)

Algebra Content Analysis		Common Core State Standards - High School Algebra
ACA 4 Graphing Linear Equations & Functions	ACA 4.1 Identify a line on a graph Which line on the graph is y = 2?	**Creating Equations** • Create equations that describe numbers or relationships **Reasoning with Equations and Inequalities** • Understand solving equations as a process of reasoning and explain the reasoning • Solve equations and inequalities in one variable • Solve systems of equations • Represent and solve equations and inequalities graphically
	ACA 4.2 Find the slope of a line through 2 points Find the slope of a line through (1, 3), (2, 5)	**Creating Equations** • Create equations that describe numbers or relationships Reasoning with Equations and Inequalities • Represent and solve equations and inequalities graphically
ACA 5 Writing Linear Equations	ACA 5.1 Slope-intercept form Write the equation in slope-intercept form: $m = \frac{1}{2}$ b = 3	**Creating Equations** • Create equations that describe numbers or relationships

Table 8 (Continued)

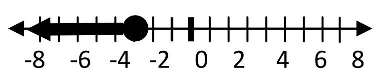
Algebra Content Analysis		Common Core State Standards - High School Algebra
		Reasoning with Equations and Inequalities • Represent and solve equations and inequalities graphically
	ACA 5.2 Write equation for line through 2 points Alternate between point-slope (ACA 5.2a) and slope-intercept form (ACA 5.2b) ACA 5.2a Write the equation of a line through (5, 3) (4, 9). Use point-slope form. ACA 5.2b Write the equation of a line through (4, 2) (6, 3). Use slope-intercept form.	**Creating Equations** • Create equations that describe numbers or relationships **Reasoning with Equations and Inequalities** • Represent and solve equations and inequalities graphically
ACA 6 Solving & Graphing Linear Inequalities	ACA 6 Interpret a graph of an inequality This graph shows the solution for which equation? 	**Reasoning with Equations and Inequalities** • Represent and solve equations and inequalities graphically
ACA 7 Systems of Linear Equations & Inequalities	ACA 7.1 Solve linear system by substitution Solve the linear system: $x - y = 4$ $x + 2y = 19$	**Reasoning with Equations and Inequalities** • Understand solving equations as a process of reasoning and explain the reasoning • Solve systems of equations

Table 8 (Continued)

Algebra Content Analysis		Common Core State Standards - High School Algebra
	ACA 7.2 Solve linear system by linear combination Solve the linear system: – 6x + 3y = – 6 2x + 6y = 30	**Reasoning with Equations and Inequalities** • Understand solving equations as a process of reasoning and explain the reasoning • Solve systems of equations
ACA 8 Exponents & Exponential Functions	ACA 8.1 Evaluate expressions with exponents (either a negative base or a negative exponent - second or third power) Evaluate the expression: 4–2	
	ACA 8.2 Simplify expressions with exponents Simplify the expression: $\frac{a^2}{ab^3} \bullet \frac{b^4}{a^3}$	**Seeing Structure in Expressions** • Interpret the structure of expressions • Write expressions in equivalent forms to solve problems

Table 9

Algebra Basic Skills skills and subskills alignment with the high school algebra Common Core State Standards

Algebra Basic Skills		Common Core State Standards-High School Algebra
ABS 1 Solving simple equations (Basic facts)	ABS 1.1 Solve simple addition equations 6 + p = 11 ABS 1.2 Solve simple subtraction equations 12 – e = 4 ABS 1.3 Solve simple multiplication equations 4r = 28 ABS 1.4 Solve simple division equations 63 ÷ c = 9	**Reasoning with Equations and Inequalities** • Understand solving equations as a process of reasoning and explain the reasoning • Solve equations and inequalities in one variable
ABS 2 Applying the distributive property	ABS 2.1 Apply the distributive property and add or subtract an integer 4(3 + s) – 7 ABS 2.2 Apply the distributive property and add or subtract a variable 5(b – 3) – b ABS 2.3 Apply the distributive property and add or subtract an integer and a variable 5(3 + f) – 2f + 6 ABS 2.4 Apply the distributive property to get a quadratic 2 + w(w – 5)	**Seeing Structure in Expressions** • Interpret the structure of expressions • Write expressions in equivalent forms to solve problems **Arithmetic with Polynomials and Rational Expressions** • Perform arithmetic operations on polynomials • Rewrite rational expressions

Table 9 (Continued)

Algebra Basic Skills		Common Core State Standards-High School Algebra
ABS 3 Working with integers	– 5 + 6 – 6 OR 9 + (– 3) – 8 (all only have 3 integers)	None of the CCSS domains align with this subskill
ABS 4 Combining like terms	ABS 4.1 Add and/or subtract linear terms with the same variable b + b + 2b ABS 4.2 Add and subtract linear variables and integers 3z – 8z + 2 + 9 ABS 4.3 Add and subtract quadratics and integers – 3w2 + 5w2 – 5 + 12 ABS 4. 4 Add and subtract linear and quadratic terms y2 + y – 4y + 3y2 ABS 4.5 Add and subtract linear terms, quadratic terms, and integers 6a + 2a – 9 + 3a2	**Seeing Structure in Expressions** • Interpret the structure of expressions • Write expressions in equivalent forms to solve problems **Arithmetic with Polynomials and Rational Expressions** • Perform arithmetic operations on polynomials • Rewrite rational expressions
ABS 5 Applying Proportional Reasoning	ABS 5.1 Calculate equivalent fractions $\frac{r}{6} = \frac{12}{18}$ ABS 5.2 Make measurement conversions 4 quarts = 1 gallon ____ quarts = 3 ¼ gallons	None of the CCSS domains align with this subskill

Research Question 2

Research question two investigates the two APMMs at the item level, including item difficulty and item discrimination. Items in each APMM are created to test specific algebraic skills and subskills. This research question investigates the levels of item difficulty that are represented in the skill/subskill categories of ACA and ABS that correspond to the Common Core State Standards for high school Algebra It also investigates the extent to which the items are able to discriminate the ability of the students.

For this research question, I arranged the results by probe type and, within each probe type, I arranged the results of the item difficulty analyses by skill/subskill categories and by CCSS high school algebra domains followed by item discrimination values. Before the item discrimination data are considered, it is important to first have a general understanding of the data. Table 10 shows the descriptive data for scores on ABS and ACA.

Table 10
Descriptive data for Algebra Basic Skills and Algebra Content Analysis probes

	N	Range	Mean	Standard Deviation
Algebra Basic Skills	35	2 - 22	11	5.60
Algebra Content Analysis	134	12 - 48	20.17	11.09

Algebra Content Analysis. In the scoring procedures used for ACA, partial credit was given for solutions in which part of the response was correct. Also, a score of negative one was awarded if a student chose the wrong answer option and did not show any work. To find the difficulty levels for items with partial credit, I used the reasoning employed by the Laboratory of Educational Research Test Analysis Package (“Item difficulty calculations”, n.d., para. 2), which is a computer program used to process and analyze results from tests and surveys. I discussed the process in detail in the Method chapter. Table 11 shows the levels of item difficulty by skill/subskill categories for ACA. There were two items in the ACA 1 and ACA 2.2 categories. To determine the level of difficulty for these two categories, the average of the difficulty levels for the items in each of these categories was calculated. The levels of difficulty across ACA subskills ranged from .30 to .91. The subskill category ‘Interpret a graph of an inequality’ had the lowest score (.30, indicating greater difficulty for students) and the subskill category ‘Evaluate expressions that include exponents and order of operations with given values’ had the highest score (.91, indicating the majority of students got it correct). Items for subskills ACA 1, ACA 2.1, ACA2.2, ACA 3.1, ACA 3.2, ACA 4.2, ACA 5.1 and ACA 5.2 had difficulty levels that indicated the items were easy (greater than .80). The remaining items in ACA had difficulty levels that indicated that they were of average difficulty (.30 - .80).

I also wanted to investigate the difficulty levels of items within the four CCSS domains for high school algebra and for that purpose I aligned all the skill/subskill categories of ACA with the four CCSS domains and conducted the difficulty analyses within these subskill groupings. Table 12 shows the difficulty levels of items for the ACA subskills within each CCSS domain. The difficulty levels ranged from .65 to .86. The items in CCSS 2 are very easy whereas the difficulty levels for items in CCSS 4 range from moderate to easy and their average of .64 indicate that the items as a group are well balanced.

Table 11

Levels of item difficulty in Algebra Content Analysis probes by skill/sub skill categories

Skills Tested	Subskills	Number of items	Number of students	Level of difficulty	Item discrimination
ACA 1 Connections to Algebra	ACA 1 Evaluate expressions that include exponents and order of operations with given alues	2	114 119	.91	53.7 46.3
ACA 2 Properties of Real Numbers	ACA 2.1 Simplify expressions that include integers and combination of like terms	1	85	.84	59.7
	ACA 2.2 Simplify expressions that include integers and combination of like terms and application of the distributive property	2	115 103	.83	44.5 58.8
ACA 3 Solving Linear Equations	ACA 3.1 Solve linear equations with 2 steps	1	114	.89	32.0
	ACA 3.2 Solve equations with variables on both sides	1	104	.86	51.5
ACA 4 Graphing Linear Equations & Functions	ACA 4.1 Identify a line on a graph	1	81	.69	52.1
	ACA 4.2 Find the slope of a line through 2 points	1	68	.84	59.2

Table 11 (Continued)

Skills Tested	Subskills	Number of items	Number of students	Level of difficulty	Item discrimination
ACA 5 Writing Linear Equations	ACA 5.1 Slope-intercept form	1	86	.86	85.4
	ACA 5.2 Write equation for line through 2 points	1	65	.80	59.2
ACA 6 Solving & Graphing Linear Inequalities	ACA 6 Interpret a graph of an inequality	1	69	.30	31
ACA 7 Systems of Linear Equations & Inequalities	ACA 7.1 Solve linear system by substitution	1	54	.72	49.9
	ACA 7.2 Solve linear system by linear combination	1	56	.75	56.9
ACA 8 Exponents & Exponential Functions	ACA 8.1 Evaluate expressions with exponents	1	102	.35	47.5
	ACA 8.2 Simplify expressions with exponents	1	68	.57	50

Table 12

Levels of item difficulty in Algebra Content Analysis by Common Core State Standards Domains

CCSS	Skill categories in ACA	Difficulty level N = 134	Average difficulty by CCSS
1. Seeing Structure in Expressions	ACA 1	.91	.79
	ACA 2.1	.84	
	ACA 2.2	.83	
	ACA 8.2	.57	
2. Arithmetic with Polynomials and Rational Expressions	ACA 1	.91	.86
	ACA 2.1	.84	
	ACA 2.2	.83	
3. Creating Equations	ACA 4.1	.69	.80
	ACA 4.2	.84	
	ACA 5.1	.86	
	ACA 5.2	.80	
4. Reasoning with Equations and Inequalities	ACA 3.1	.89	.65
	ACA 3.2	.86	
	ACA 4.1	.69	
	ACA 4.2	.84	
	ACA 5.1	.86	
	ACA 5.2	.80	
	ACA 6	.30	
	ACA 7.1	.72	
	ACA 7.2	.75	

The item discrimination levels for ACA are reported in the last column of Table 11. Thirteen items in ACA had discrimination levels ranging from 46.3 to 85.4 indicating that they were excellent items. Scores on the remaining three items ranged from 31 to 34.5 indicating that they were good items.

Algebra Basic Skills. In ABS, a score of 1 was given for a correct response and 0 for an incorrect response. Unlike ACA, where most subskills are represented by just one item, each subskill in ABS has multiple items that are equivalent. In all, there are 60 items in ABS, which in turn are distributed within sixteen subskill categories, the details of which were discussed in the measures section of the Method chapter. If a student attempted at least one item successfully within a subskill category, that student's response was coded as being correct and a value of one was assigned for that subskill. Otherwise (e.g., none of the items for the subskill were answered correctly), the subskill was coded as incorrect and a value of zero was assigned for that subskill. Table 13 shows levels of item difficulty by skill/subskill categories for ABS. The difficulty levels in ABS subskills ranged from .03 to .83, with ABS 4.3 and 4.4 having values of .03, indicating that these items were very difficult. From Table 13, it can be seen that each of these subskills has only one item. Also it was found that only three (for ABS 4.3) to six (for ABS 4.4) percent of students attempted these items as compared to ABS 2.4 and 4.5 which also have only one item but were attempted by fifty seven and twenty percent of students, respectively, and had average difficulty levels.

As with ACA, I wanted to investigate the difficulty levels within the four CCSS domains for high school algebra and for that purpose I aligned the skill/subskill categories of ABS with the four CCSS Domains and conducted the analyses for these subskill groupings. As can be seen from Table 14, such alignment reveals that CCSS 1 and CCSS 2 have the same ABS subskills, whereas none of the ABS subskills align with CCSS 3. When items are grouped by CCSS, mean levels of difficulty for each of the three CCSS are average.

Table 13

Levels of item difficulty in Algebra Basic Skills probes by skill/sub skill categories

Skills Tested	Subskills	Number of items per subskill	Levels of difficulty N= 35
ABS 1 Solving simple equations (Basic facts)	ABS 1.1 Solve simple addition equations	3	.63
	ABS 1.2 Solve simple subtraction equations	3	.83
	ABS 1.3 Solve simple multiplication equations	3	.60
	ABS 1.4 Solve simple division equations	3	.43
ABS 2 Applying the distributive property	ABS 2.1 Apply the distributive property and add or subtract an integer	5	.26
	ABS 2.2 Apply the distributive property and add or subtract a variable	5	.20
	ABS 2.3 Apply the distributive property and add or subtract an integer and a variable	9	.37
	ABS 2.4 Apply the distributive property to get a quadratic	1	.43
ABS 3 Working with integers		10	.71

Table 13 (Continued)

Skills Tested	Subskills	Number of items per subskill	Levels of difficulty N= 35
ABS 4 Combining like terms	ABS 4.1 Add and/or subtract linear terms with the same variable	3	.23
	ABS 4.2 Add and subtract linear variables and integers	4	.26
	ABS 4.3 Add and subtract quadratics and integers	1	.03
	ABS 4. 4 Add and subtract linear and quadratic terms	1	.03
	ABS 4.5 Add and subtract linear terms, quadratic terms, and integers	1	.20
ABS 5 Applying Proportional Reasoning	ABS 5.1 Calculate equivalent fractions	4	.40
	ABS 5.2 Make measurement conversions	4	.49

As explained in the method chapter, the nature of Algebra Progress Monitoring measures is such that students usually cannot attempt all the items in the test within the stipulated time. Hence, in the sample of 35 students for ABS, there were more items left unattempted by the students than attempted rendering the sample insufficient to run analysis and get valid results. As a result, discrimination statistics were not computed for the ABS measure.

Table 14

Levels of item difficulty in Algebra Basic Skills by Common Core State Standards Domains

Common Core State Standards Domains	Skill categories in ABS	Difficulty level N = 35	Average difficulty by CCSS
Seeing Structure in Expressions	ABS 2 .1	.26	.22
	ABS 2 .2	.20	
	ABS 2 .3	.37	
	ABS 2 .4	.43	
	ABS 4 .1	.23	
	ABS 4 .2	.26	
	ABS 4 .3	.03	
	ABS 4 .4	.03	
	ABS 4 .5	.20	
Arithmetic with Polynomials and Rational Expressions	ABS 2 .1	.26	.22
	ABS 2 .2	.20	
	ABS 2 .3	.37	
	ABS 2 .4	.43	
	ABS 4 .1	23	
	ABS 4 .2	.26	
	ABS 4 .3	.03	
	ABS 4 .4	.03	
	ABS 4 .5	.20	
Creating Equations			
Reasoning with Equations and Inequalities	ABS 1.1	.63	.62
	ABS 1.2	.83	
	ABS 1.3	.60	
	ABS 1.4	.43	

Research Question 3

The third research question deals with the predictive power of ABS and ACA in predicting student performance in state achievement tests, in this case, ITBS and ITED. The earlier technical reports (Perkmen, Foegen, & Olson, 2006; Foegen & Olson, 2007a; Foegen & Olson, 2007b) indicated that the predictive validity for total scores in both these measures to predict student performance in ITBS and ITED were not very encouraging, with correlations in the .07 to .27 range. This study extended that investigation by examining the extent to which the subtotal scores, based on the skill/subskill categorization of the measures, predict performance on state achievement tests in comparison to total scores. The study also investigated the predictive power of subtotal scores when the items were categorized by CCSS.

To examine the predictive validity of the total scores and subscores of ACA and ABS, I computed the correlations between students' scores on probe subskills and their ITBS or ITED math scores. Students in grades 9 to 12 completed the ITED, while students in grade 8 took the ITBS. District records were used to access students' scores on these instruments; national percentile ranks were used for the analyses.

Algebra Content Analysis. Table 15 reports the correlations between the total scores and subskill scores on each probe with scores on the ITED and ITBS tests. There was only one strong significant relationship within subskill categories and state achievement tests; that was between ACA 3.1 and ITBS math total (r = .73). Relationships between the remaining subskills and ITBS Math Total scores were not significant. The relation between the ACA total score and ITBS math total (r = .48) was moderate and although relationships with ACA total scores were significant for ITED Computation (r = .23) and ITED Problems/Data Interpretation (r = .27), they were weak. Relations between the ACA subskills and ITED Computation were statistically significant but weak for ACA 1 (r = .28) and ACA 3.1 (r = .20). This indicates that there is significant (though weak) predictive power for ITED Computation in the subskill categories 'evaluating expressions' and 'solving linear equations in two steps.' The remaining subskill categories had no significant relations with ITED Computation scores. Relations between ACA subskills and ITED Concepts/Problem Solving were not significant except for ACA 2.2, that is 'simplify expressions' (r = .22).

Table 15
Correlations between ACA subskills and ITED and ITBS scores

		ITED Computation N=104	ITED Con/ Prob N=104	ITBS ConEst N=21	ITBS Math Total N=20
Total Score	Pearson Correlation	.23*	.27**	.33	.48*
	Sig. (2-tailed)	.017	.005	.142	.032
ACA1	Pearson Correlation	.28**	.16	.24	.15
	Sig. (2-tailed)	.003	.104	.297	.537
ACA 2.1	Pearson Correlation	.14	.15	-.13	-.11
	Sig. (2-tailed)	.145	.120	.578	.642
ACA2.2	Pearson Correlation	.08	.22*	.26	.42
	Sig. (2-tailed)	.404	.023	.247	.066

*Correlation is significant at the 0.05 level (2-tailed).
**Correlation is significant at the 0.01 level (2-tailed).
Note. ITED = Iowa Test of Educational Development; Con/ Prob = Concepts/problem Solving; ITBS = Iowa Tests of Basic Skills; ConEst = Concepts and Estimation

Table 15 (Continued)

		ITED Computation N=104	ITED Con/ Prob N=104	ITBS ConEst N=21	ITBS Math Total N=20
ACA 3.1	Pearson Correlation	.20*	.07	.36	.73**
	Sig. (2-tailed)	.037	.515	.109	.000
ACA 3.2	Pearson Correlation	.15	.12	-.13	.05
	Sig. (2-tailed)	.122	.217	.579	.850
ACA 4.1	Pearson Correlation	.09	.19	.36	.44
	Sig. (2-tailed)	.365	.058	.110	.054
ACA 4.2	Pearson Correlation	.04	.000	.07	.31
	Sig. (2-tailed)	.720	.998	.776	.186
ACA 5.1	Pearson Correlation	.12	.19	-.02	.00
	Sig. (2-tailed)	.240	.057	.922	.985
ACA 5.2	Pearson Correlation	.13	.00	.00	-.01
	Sig. (2-tailed)	.187	.985	.993	.960
ACA 6	Pearson Correlation	.14	.09	.04	.16
	Sig. (2-tailed)	.168	.370	.874	.506
ACA 7.1	Pearson Correlation	-.00	.13	.28	.34
	Sig. (2-tailed)	.988	.206	.227	.139

Table 15 (Continued)

		ITED Computation N=104	ITED Con/ Prob N=104	ITBS ConEst N=21	ITBS Math Total N=20
ACA 7.1	Pearson Correlation	.18	.14	.33	.34
	Sig. (2-tailed)	.075	.152	.142	.146
ACA 8.1	Pearson Correlation	-.09	.12	.22	-.07
	Sig. (2-tailed)	.362	.244	.343	.760
ACA 8.2	Pearson Correlation	.02	.16	.19	.25
	Sig. (2-tailed)	.841	.114	.414	.283

Table 16 shows the correlations between the probe subskills (arranged within the CCSS domains) and ITED and ITBS scores. The relationship between ITBS and CCSS 4 (r = .46) was the only significant relationship that had moderate strength. ITED Computation and ITED Concept/Problem Solving (Con/Prob) scores had significant but weak relationships with CCSS1, 2 and 4. CCSS 3 did not have significant relations with ITED or ITBS score.

Algebra Basic Skills. Table 17 shows the correlations between ABS total scores, scores for ABS subskills and ITED scores. Relations between ABS 2.3 (r = .40), ABS 2.4 (r = .36), ABS 3(r = .45), and ABS 5.1 (r = .36) and ITED Computation were moderate. In other words, applying distributing property, working with integers, and calculating equivalent fractions were the subskills that had moderate relationships with ITED computation. The remaining subskills did not show significant relationships with ITED Computation. There were no significant relations between ABS scores and ITED Concept/Problem Solving (Con/Prob) scores. There was no significant relation between ABS total scores and ITED scores. ABS subskills, when grouped by CCSS for high school algebra, showed no significant relationship with ITED.

Table 16

Correlations between ACA subskills grouped within Common Core State Standards for high school and ITED and ITBS scores

		ITED Computation N=104	ITED Con/Prob N=104	ITBS ConEst N=21	ITBS Math Total N=20
CCSS1	Pearson Correlation	.22*	.26**	.15	.27
	Sig. (2-tailed)	.027	.007	.519	.242
CCSS2	Pearson Correlation	.25*	.26**	.13	.25
	Sig. (2-tailed)	.011	.008	.583	.296
CCSS3	Pearson Correlation	.14	.15	.20	.31
	Sig. (2-tailed)	.153	.126	.395	.180
CCSS4	Pearson Correlation	.23*	.21*	.28	.46*
	Sig. (2-tailed)	.020	.034	.227	.042

*Correlation is significant at the 0.05 level (2-tailed).
**Correlation is significant at the 0.01 level (2-tailed).
Note. CCSS = Common Core State Standards; ITED = Iowa Test of Educational Development; Con/ Prob = Concepts/problem Solving; ITBS = Iowa Tests of Basic Skills; ConEst= Concepts and estimation

Table 17

Correlations between ABS subskills and ITED scores

		ITED Con/ Prob N=31	ITED Computation N=31
Total Score	Pearson Correlation	.07	.27
	Sig. (2-tailed)	.701	.139
ABS 1.1	Pearson Correlation	-.09	-.16
	Sig. (2-tailed)	.643	.404
ABS 1.2	Pearson Correlation	.08	.14
	Sig. (2-tailed)	.655	.461
ABS 1.3	Pearson Correlation	-.16	.06
	Sig. (2-tailed)	.382	.755
ABS 1.4	Pearson Correlation	.06	-.17
	Sig. (2-tailed)	.745	.356
ABS 2.1	Pearson Correlation	.12	.11
	Sig. (2-tailed)	.507	.565
ABS 2.2	Pearson Correlation	.15	.23
	Sig. (2-tailed)	.430	.224

*Correlation is significant at the 0.05 level (2-tailed).
Note. ITED = Iowa Test of Educational Development; Con/ Prob = Concepts/problem Solving

Table 17 (Continued)

		ITED Con/ Prob N=31	ITED Computation N=31
ABS 2.3	Pearson Correlation	.22	.40*
	Sig. (2-tailed)	.229	.025
ABS 2.4	Pearson Correlation	.17	.36*
	Sig. (2-tailed)	.372	.048
ABS 3	Pearson Correlation	.26	.45*
	Sig. (2-tailed)	.164	.011
ABS 4.1	Pearson Correlation	.18	.16
	Sig. (2-tailed)	.336	.396
ABS 4.2	Pearson Correlation	.19	-.04
	Sig. (2-tailed)	.313	.852
ABS 4.3	Pearson Correlation	-.11	.19
	Sig. (2-tailed)	.542	.309
ABS 4.4	Pearson Correlation	-.11	.19
	Sig. (2-tailed)	.542	.309
ABS 4.5	Pearson Correlation	-.01	.03
	Sig. (2-tailed)	.954	.883
ABS 5.1	Pearson Correlation	.15	.36*
	Sig. (2-tailed)	.421	.048

Table 17 (Continued)

		ITED Con/ Prob N=31	ITED Computation N=31
ABS 5.2	Pearson Correlation	-.07	-.19
	Sig. (2-tailed)	.716	.300

Table 18

Correlations between ABS subskills grouped within Common Core State Standards for high school and ITED and ITBS scores

		ITEDConProbNPR N=31	ITEDCompNPR N=31
CCSS1and2	Pearson Correlation	.20	.32
	Sig. (2-tailed)	.271	.081
CCSS4	Pearson Correlation	-.05	-.06
	Sig. (2-tailed)	.805	.755

Note. ITED = Iowa Test of Educational Development; Con/ Prob = Concepts/problem Solving; Comp= = Computation; NPR = National percentile rank

Summary of Results

The results above indicate that the content of ACA is a good fit with CCSS for high school algebra. With regard to ABS there is a lot of overlap and exclusion when the content being tested in ABS is aligned with the CCSS for high school algebra. When investigating the difficulty levels of the measures, the items in both ABS and ACA are mostly found in the ‘average difficulty’ level though some ABS items have very high difficulty level. Items in ACA have also displayed good discriminating power in terms of student ability. The subtotals for subskill categories in ACA and ABS were investigated for their predictive validity. All these results are discussed in detail in the next chapter.

Chapter 5: Discussion

Discussion of Results

With the recent adoption of the Common Core State Standards (CCSS) by the state of Iowa (along with forty one other states), it is important to know how well the content tested in the existing algebra progress monitoring measures aligns with these standards. This study investigated the extent to which the content in Algebra Basic Skills (ABS) and Algebra Content Analysis (ACA) aligns with the CCSS. In addition, items in ABS and ACA were investigated with respect to classical item analysis, including item difficulty and item discrimination. The predictive validity of ABS and ACA total scores and subskill scores were also investigated against ITED and ITBS. The discussion of the results is organized using the research questions.

Research question 1: Content validity with respect to CCSS. Algebra is often considered the gateway to higher mathematics; as a result, a lot of emphasis is given to instruction in Algebra. At the same time it is important to track student learning. Given the high stakes associated with Annual Yearly Progress goals, it is important for schools to know where their students stand in terms of results generated by state achievement tests. This raises the issue of monitoring student learning throughout the school year and hence the importance of tests. A reliable and valid test is an important tool that assists teachers by informing them about the quality of their instruction and the extent of student learning. It is very important that a test is assessing the content being taught. With the state of Iowa adopting the CCSS, the mathematics content currently being taught in schools in Iowa will need to be modified according to the recommendations in the CCSS. It will then be important to look at the tests used in the past by schools to see if they continue to test what is being taught in class.

Algebra Progress Monitoring Measures were developed before the CCSS came into existence. These measures were constructed based on the algebra content in traditional textbooks and teacher feedback (Foegen & Lind, 2004). So for these measures to provide valid results, it is important to investigate the relationship between the content tested in these measures and that prescribed in CCSS.

The alignment of content in ABS and ACA with the domains in CCSS-high school Algebra showed that ACA is better aligned with the standards than ABS. However, the standards for high school algebra encompass all the algebra content expected to be taught in high school, whereas Algebra Progress Monitoring Measures were developed for Pre Algebra and Algebra 1, a small component of high school algebra. At the present time, however, Algebra is generally taught as a formal course in high school and the probes are administered in high school. Given this scenario, I chose to align ABS and ACA with the high school algebra standards within CCSS. As a result, there are subskills in the measures that do not belong to any particular standard and there are subskills that belong in more than one standard. On the other hand, there are standards that do align with any subskills, indicating that that standard has not been covered in the measures.

A look at the subskills that did not belong in any of the four CCSS reveals that they are categories that basically require arithmetic skills involving integers, fractions, and exponents. For example ACA 8.1 requires solving exponents, ABS 3 is about working with integers, and ABS 5.1 is calculating equivalent fractions. When the alignment is considered from the perspective of the CCSS domains, I observed that the ACA subskills were well distributed among the four CCSS domains. In terms of ABS, I observed that this distribution was not even. CCSS 1 and CCSS 2 have the same subskill categories whereas CCSS 3 did not have any subskill categories indicating that ABS does not cover all the content prescribed by CCSS. If we do a cross sectional observation of subskills and CCSS, there are two subskills that do not belong to any of CCSS domains and one CCSS domain that does not have any subskill categories. This uneven distribution and the exclusions can be explained by the fact that ABS mostly caters to early or pre algebra skills and hence much of the more advanced algebra content is missing. Interestingly, the missing components (in both ABS and ACA) fit very nicely into the Grade 8 CCSS domain of ‘Expressions and Equations’ which has the following clusters:

1. Work with radicals and integer exponents
2. Understand the connections between proportional relationships, lines, and linear equations.
3. Analyze and solve linear equations and pairs of simultaneous linear equations.

This result suggests that the ABS content is more aligned with Grade 8 CCSS for mathematics than with high school algebra. As a result, eighth grade teachers may wish to use the algebra measures to monitor progress. They must keep in mind that the algebra measures reflect only one Domain (there are five CCSS Domains in grade 8 mathematics) and therefore the measure’s utility will be limited to a part of the curriculum. As it happens, progress monitoring measures are not intended to measure mastery, but to monitor progress across a particular interval of time (mostly through an academic year or a semester). Aligning ABS content with Grade 8 CCSS for mathematics would result in considering only a part of the content the students are being taught in a whole school year as compared to high school where algebra is taught the whole school year or semester. The use of ABS measure makes more sense when algebra is taught as a course rather than part of a course. Because of this, I considered the CCSS for high school algebra for my study rather than the standards for Grade 8.

Research question 2: Item difficulty and discrimination. My second research question explored classical item analysis involving item difficulty and item discrimination for both ABS and ACA. For item difficulty, items with difficulty levels ranging from .20 to .90 are considered good (Allen & Yen, 1979). Such items would have average difficulty and would be ideal. Items with difficulty levels above .90 are considered to be very easy and those below .20 are considered to be very difficult and are not preferred. The results of this study indicated that for ACA, only ACA 1, with difficulty level .91, was above the preferred range of difficulty levels. The rest of the items had average difficulty levels. For ABS, there were two item categories for which the results indicated they were very difficult. Items in ABS 4.3 (Add and subtract quadratics and integers) and 4.4 (Add and subtract linear and quadratic terms) produced difficulty levels of .03. In fact, the whole ABS 4 with its five subskill categories produced difficulty levels that indicate high difficulty levels (.20 - .26, not including ABS 4.3 and 4.4). A comparison of the difficulty levels between ACA and ABS reveal that ACA items have values that indicate higher success rates than ABS items.

These results must be considered in context when discussing the measures' items. For the ABS difficulty levels, as described in Method chapter, I grouped all the items by subskills and coded students to be successful if even one item in that group was correct. There is a possibility that the students might have had more hits on certain subskill categories because of a greater number of items in them as opposed to some categories that have just one item, resulting in having a higher probability of success in one category versus another. For ACA, I followed LERTAP's method of finding the difficulty level. According to LERTAP's proportional method, to find item difficulty, any response is counted as being correct if its corresponding weight is greater than zero. This method does not take into account any differences that may exist among response weights. In ACA partial scores are given for items that are partially correct. So all the items that had scores above zero (fully and partially correct) were considered as correct. This might be the reason for higher difficulty scores in ACA versus for ABS where only correct responses were scored. But overall, barring a couple of items in each measure, most of the items were indicated to be good with respect to difficulty levels.

Unlike difficulty levels, item discrimination for both measures was done item wise. Items in ACA produced discrimination values indicating excellent and good discriminating power. Items in ACA 1, ACA 2, ACA 3.2, ACA 4, ACA 5, ACA 7, and ACA 8 indicated that they have excellent discriminating powers (44.5 - 85.4). Items in ACA 3.1 and ACA 6 indicate that they have good discriminating powers (31- 32). In the case of ABS, due to the large number of items in this measure, there were more unattempted items than items that were attempted. In addition, the sample size for ABS was 35 and with a large number of unattempted items, I did not have sufficient numbers of attempts per item to be able to determine item discrimination. As a result, I decided not to include these analyses in this study.

With respect to difficulty levels, in ACA the difficulty levels were towards the higher end of the success rate. If the difficulty levels of items are compared with their discrimination values, it shows that even though the items have high difficulty statistics (indicating that they are easy), they still have good discriminating powers.

Research question 3: Predictive validity. The third research question deals with the predictive powers of ABS and ACA for predicting performance in ITED and ITBS. The predictive validity of the measures was examined by correlating scores on the measures with scores in ITED and ITBS. Predictive validity for total scores on ABS and ACA have already been investigated in technical reports for Algebra Progress Monitoring Measures. This study extended this investigation to the subskill categories level. As described in the earlier chapters, I was interested in finding out whether any of the subskills in these measures would have a significant relation in predicting student performance in ITED or ITBS. I was also interested in how the relations between subskills and state achievement tests compared to relations between total scores on the progress monitoring measures and achievement tests.

For ACA, previous studies (Foegen & Olson, 2007; Perkmen et al., 2006a; 2006b; 2006c) showed correlations with ITED Computation ranging from .30 to .64 and correlations with ITED Con/ Prob. ranging from .25 to .56. Correlations between ACA and ITBS Prob/Data ranged from .32- .34 (Foegen & Olson, 2007). For the present study, correlations were near the lower end of these ranges with the exception of ITBS Prob/Data which had a non significant relation. This exception might have occurred due to the lower sample size (N = 21) in the current study.

This study extended the investigation of the predictive powers of ACA by exploring the predictive validity of scores in ACA subskills with scores in ITED and ITBS. Not many statistically significant results emerged from this inquiry. ACA 1, which is ‘Evaluate expressions that include exponents and order of operations with given values,’ had a significant relation at the .01 level (r = .28) with ITED Computation; ACA 3.1, which is ‘Solve linear equations with two steps,’ also had a significant relation at .05 level (r = .20) with ITED Computation. ACA 2.2, which is ‘Simplify expressions that include integers and combination of like terms and application of the distributive property,’ had a significant relation at the .05 level (*r* = .22) with ITED Prob/Data. Correlations for the ACA subskills with ITBS showed a significant relation with only one subskill. ACA 3.1 had a strong relation with ITBS at the .01 level (*r* = .73). This indicates that the skill of solving simple equations is correlated to ITED Computation and ITBS content.

For ABS total scores, the correlations obtained in previous studies with ITED Computation ranged from .33 to .47, correlations with ITED Con/ Est. ranged from nonsignificant to .45, and correlations with ITED Prob/Data were not significant. In the present study, correlations between ABS total scores and ITED Computation and Con/Prob were not significant. This disparity in the results between earlier studies and the current study might be due to the limited sample size used in this study (N = 31). Regarding correlations between ABS subskills and ITED Computation and ITED Con/Prob, there were no significant relations with ITED Con/Prob. Significant correlations were obtained between ITED Computation and ABS 2.3 (Apply the distributive property and add or subtract an integer and a variable) and ABS 3   (Working with integers), with coefficients of .40 and .45 respectively.

As an extension to my exploration in terms of CCSS, I grouped the subskills of ACA and ABS within the four CCSS Domains. I found that this arrangement had significant relations in terms of predicting scores in ITED and ITBS. In ACA, scores in CCSS 1 and CCSS 2 had a significant relation with both ITED Computation (*r* = .25 - .22) and Con/Prob (*r* = .26) and scores in CCSS 4 had significant relations with ITED Computation (*r* = .23), Con/Prob (*r* = .21), and ITBS (*r* = .46). CCSS 3, which is ‘Creating Equations’ did not have a significant relation with ITED or ITBS. CCSS 3 in ACA consists of ‘Graphing Linear Equations & Functions’ and ‘Writing Linear Equations’ skills. In ABS, none of the scores in CCSS had a significant relationship with ITED Computation or Con/Prob.

Limitations

Limitations for this study can be broadly categorized into three parts: Limitations due to data; limitations due to the construction of the measures; and last but not least, lack of earlier studies to guide this study.

Limitations due to available data. This study was conducted using data that were collected for earlier studies; as a result the sample size was not a choice for this study. Also the data used for this study were collected over a period of time (September- November 2006) as opposed to all at one time; this could have led to some discrepancies in student performances (performance of a student in September as compared to that of a student in November).

Limitations due to the design of APMM. Curriculum based measures are designed to measure students’ progress across a period of time in contrast to traditional tests that evaluate student ability at a particular time. Hence the construct of CBM is such that the students usually cannot complete all the items in the test. This is done to avoid a ceiling effect. A ceiling effect occurs when a student successfully attempts all the items in the measure and experiences a “ceiling,” or limit in maximizing his score. The growth of a student is measured by comparing the present and the previous performance of a student on a measure. In cases where the student has successfully attempted all the items in the measure in successive administrations, the results would not indicate any growth in student learning which would not be a true assessment of student growth.

Due to the design of the APMMs, all items were not attempted by the students and therefore item analysis had to be conducted using certain criteria. For ABS, items testing the same subskill were aggregated for item difficulty and recoded as one item as described in the Method chapter. Moreover all the subskills in ABS do not have uniform numbers of items resulting in a range of probabilities for attempting each subskill. For example, ABS 4.3 has one item, ABS 1.1 has 3 items, and ABS 3 has 10 items. One successful attempt in a subskill was considered a success in that subskill which would mean that a student has three chances for success in ABS 1.1 and ten chances for success in ABS 3 as compared to ABS 4.3 where a student has only one chance to succeed. Though it is limited, this discrepancy also exists in ACA where only two subskills have different numbers of items (two per subskill) in contrast to the rest of the subskills (which have one item per subskill). The effect of this discrepancy cannot be ignored when considering the results for predictive validity, where subtotal scores were considered based on the subskill categorization and later when these subskill categories were further grouped by CCSS for high school algebra.

Due to the large number of items in ABS and a limited time of five minutes to attempt the test, many items in ABS were left unattempted. This resulted in a very small sample of attempted items which was not large enough to run analyses for item discrimination.

Limitations due to lack of earlier studies to guide this study. There are no studies preceding this study that addressed the first two research questions. Content validity for standardized tests and CBM have been established by the test authors by describing the process adopted to select test items. But there is no established process to quantify that validity except by comparing test results to other tests being used in parallel to these tests (whose content validity is in question again). But how does one examine whether a test is actually evaluating the content of the curriculum? In this context I had to develop my own criteria for establishing content validity for ABS and ACA against CCSS for high school algebra.

Investigations at the item level for CBM are very limited. I have come across two studies relating to item level analysis and both used Item Response Theory (IRT). One study was conducted on the Early Numeracy Indicators (CBM math measures for kindergarten and first grade) (Braun-Monegan, 2010**)** and the other study was conducted on ABS (Hoffman, n.d.). Whereas these studies had to establish some criteria to run their analyses to overcome the design of CBM, the analyses require skill sets that I am yet to master and so I chose to conduct classical item analysis.

Implication for Future Research

This study brought to light many new aspects about ABS and ACA and confirmed some results from the earlier studies as well as offering directions for future research. Though results showed that the content in ACA was a good fit to the CCSS, there was no way to quantify the degree of fit. It would be interesting to know the relationship between these two once the CCSS are implemented and we can have access to scores for tests aligned to the CCSS and administered in algebra classes. As noted in the discussion section, the content in ABS has a better fit with the grade 8 CCSS and may be used for the algebra part of the curriculum.

Item analysis for ABS was not satisfactory because of the large number of items in this measure, which resulted in many items left unattempted by the students. In the future, a study could be undertaken where students attempt this measure with added time (standard administration procedures for the test include a duration of five minutes) or in an untimed format so students can attempt most of the items. This would enable researchers to get a good sample of attempted items and item analysis can be properly done.

Predictive validity for both ABS and ACA were not particularly encouraging. This may be due to the fact that there is not much of the algebra content in ITBS and ITED whereas both ABS and ACA measure pre algebra and algebraic skills. However, one can test the predictive validity of these measures with the Iowa End-of-Course Assessments (IEOC) for algebra that were developed at The University of Iowa (Iowa End- of - Course assessments, n.d.) and are standardized even though they are not yet used as high stakes tests. The results obtained will give a good idea of the validity of these measures with standardized algebra tests.

Summary

This study explored the content validity, item level difficulty levels, item discrimination levels, and predictive validity of two APMMs, ABS and ACA. The results indicated that while ACA has good alignment with CCSS for high school algebra because ABS is more directed towards pre algebra, there was a lot of overlap and exclusion of the CCSS when aligned with ABS subskills. Standards in Domain 3 for high school algebra did not align with any of the ABS subskills and Domains 1 and 2 were identical having alignment with ABS 2 and ABS 4.

Item level analyses indicated that the items in ACA had a broad range of difficulty levels ranging from .3 to .9 and were comparable to the recommended levels of difficulty from .2 to .8, which provide the maximum information about difficulty among students. The difficulty levels in ABS subskills which ranged from .03 to .83 also fall in this recommended difficulty range and are thus good items in terms of difficulty. All the items in ACA produced good discrimination levels indicating that they were good in discriminating students based on their competence in algebraic skills. Item discrimination analysis for ABS was not performed due to an inadequate sample of attempted items.

Predictive validity results for ACA reiterated the results from earlier studies that were conducted to establish technical adequacy for the APMMs. Predictive validity for ACA was moderate with ITBS Math Total scores and weak with ITED Computation scores. Subskill predictive validity for ACA with ITBS was the strongest for ACA 3.1 (solve linear equations) as compared to any other subskills with ITBS/ITED scores. Predictive validity was not significant for the other subskills. Subskill predictive validity for ACA with ITED was weak for ACA 1 and ACA 3 and not significant for the other subskills. When grouped by the CCSS Domains, CCSS 4 (Reasoning with Equations and Inequalities) had a weak relation with ITBS and CCSS 1 (Seeing Structure in Expressions), CCSS 2 (Arithmetic with Polynomials and Rational Expressions), and CCSS 4 (Reasoning with Equations and Inequalities) had weak relationships with ITED. CCSS 3 (Creating Equations) did not have significant relationships with either ITED or ITBS.

Predictive validity results for ABS with ITED indicated no significant results when total scores were considered. Subskill wise, ABS 2.3 and 2.4 (apply distributive property), ABS 3 (working with integers), and ABS 5.1 (calculate equivalent fractions) had moderate relationships with ITED Computation scores. All in all, there was no pattern established that supported the use of total score or subskill scores to effectively predict student performance for ITBS or ITED.

This study which was exploratory in many aspects indicated that the content in ABS and ACA was comparable to the CCSS. The item difficulty levels in both measures indicated that they were good to test a broad range of student abilities. Item discrimination levels indicated that items in ACA had good discriminating powers. Predictive validity resultsfor ACA total scores indicated weak but significant relationship between ACA total scores and ITBS /ITED scores. Similar results were obtained when ACA subskills were grouped by the CCSS domains.The subtotals in ACA showed a weak but significant relationship with ITED Computation scores (with the exception of the ‘Creating equations’ standard). Subtotals by subskills in ACA did not show many significant relationships. Further research is need to refine the alignment of the APMMs and explore the measures’ item level characteristics in order to enhance their practical utility for teachers seeking to improve their students’ achievement in algebra.

Appendix A: IRB letter

IOWA STATE UNIVERSITY
OF SCIENCE AND TECHNOLOGY

Institutional Review Board
Office for Responsible Research
Vice President for Research
1138 Pearson Hall
Ames, Iowa 50011-2207
515 294-4566
FAX 515 294-4267

DATE: January 11, 2010

TO: Subhalakshmi Singamaneni
N131 Lagomarcino

CC: Anne M. Foegen
N162D Lagomarcino

FROM: Office for Responsible Research

TITLE: **Examining item level characteristics of two algebra progress monitoring measures**

IRB ID: 09-556

Submission Type: New **Exemption Date:** 8 January 2010

The project referenced above has undergone review by the Institutional Review Board (IRB) at Iowa State University and has been declared exempt from the requirements of the human subject protections regulations as described in 45 CFR 46.101(b). The IRB determination of exemption means that:

- **You do not need to submit an application for annual continuing review.**
- **You must carry out the research as proposed in the IRB application**, including obtaining and documenting informed consent if you have stated in your application that you will do so or if required by the IRB.
- **Any modification of this research should be submitted to the IRB on a Continuing Review and/or Modification form, prior to making any changes**, to determine if the project still meets the federal criteria for exemption. If it is determined that exemption is no longer warranted, then an IRB proposal will need to be submitted and approved before proceeding with data collection.

Please be sure to **use only the approved study materials** in your research, including the **recruitment materials and informed consent documents that have the IRB approval stamp.**

Please note that you must submit all research involving human participants for review by the IRB. **Only the IRB may make the determination of exemption**, even if you conduct a study in the future that is exactly like this study.

Appendix B: Sample Algebra Basic Skills probe

Algebra Basic Skills 1

Solve: $9 + a = 15$	$a =$
Evaluate: $12 + (-8) + 3$	
Simplify: $2x + 4 + 3x + 5$	
Solve: $12 - k = 4$	$k =$
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: $2k + 3 - 5(k + 7)$	

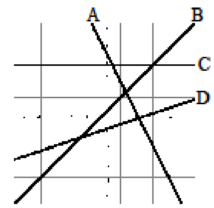
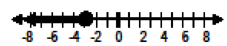
Page 1

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: $\frac{12}{2} = \frac{48}{m}$	$m =$
Simplify: $y^2 + y - 4y + 3y^2$	
Simplify: $3(c + 2) - 2c$	

OSEP Award# H324C030060

Appendix C: Sample Algebra Content Analysis probe

Algebra Content Analysis 1 — Page 1

Solve: $3x + 4 = 19$ $x =$ a) 8 b) 22 c) 15 d) 5	Evaluate $a^2 - b \div 2$ when $a = 4$ and $b = 6$ a) 1 b) 5 c) 10 d) 13	Which line on the graph is $y + 2x = 4$?  a) Line A b) Line B c) Line C d) Line D	Simplify: $3(m + 2) + 2(m - 1)$ a) $5m + 4$ b) $5m + 1$ c) $6m + 8$ d) $6m - 8$
Evaluate the expression: 6^{-2} a) -36 b) $\frac{1}{36}$ c) $\frac{1}{12}$ d) -12	Solve the linear system: $x - y = 4$ $x + 2y = 19$ a) $(-1, -5)$ b) $(5, 8)$ c) $(-2, 19)$ d) $(9, 5)$	This graph shows the solution for which inequality?  a) $x > -3$ b) $2x < -6$ c) $-3x > 9$ d) $3x > 9$	Write the equation in slope-intercept form if $m = \frac{1}{2}$ and $b = 3$ a) $y = 2x + 3$ b) $y = 3x + \frac{1}{2}$ c) $x = \frac{1}{2}y - 3$ d) $y = \frac{1}{2}x + 3$

OSEP Award# H324C030060

References

Allen, M. J., & Yen, W. M. (1979). *Introduction to measurement theory*. Monterey, California: Brookes/Cole.

Braun-Monegan, J.(February, 2010). *Item Analysis and alternate forms.* Session presented at the 17th Annual Pacific Coast Research Conference, San Diego, CA.

Common Core State Standards Initiative. (n.d). *About the standards*. Retrieved from http://www.corestandards.org/about-the-standards.

Common Core State Standards Initiative. (n.d). *The Standards: Mathematics*. Retrieved from http://www.corestandards.org/the-standards/mathematics.

Deno, S. L. (1985). Curriculum-based measurement: The emerging alternative. *Exceptional Children*, *52*, 219-232.

Deno, S. L. (2003). Developments in curriculum-based measurement. *Remedial and Special Education*, *37*, 184-192.

Deno, S. L. & Mirkin, P. K. (1977). *Data-based program modification: A manual*. Reston, VA: Council for Exceptional Children.

Findley, W.G. (1956). A rationale for evaluation of item discrimination statistics. In A. Ward, H. W. Stoker, M. Murray-Ward (Eds.), *Educational Measurement: Theories and applications* (pp. 79-84). University Press of America, Maryland.

Foegen, A., & Olson, J. R. (2005). *Reliability and criterion validity of two algebra measures: Translations and Content Analysis-Multiple Choice (Technical report 6)*. Project AAIMS, Department of Curriculum and Instruction, Iowa State University, Ames, IA.

Foegen, A., & Olson, J. (2007). *Effects of teachers' engagement with student data on students' algebra progress*. (Technical Report 16). Ames, IA: Iowa State University, Department of Curriculum and Instruction, Project AAIMS.

Foegen, A., Olson, J. R., & Impecoven-Lind, L. (2008). Developing progress monitoring measures for secondary mathematics: An illustration in algebra. *Assessment for Effective Intervention*, *33*, 240-249

Foegen, A., Olson, J., & Perkmen, S. (2005). *Reliability and criterion validity of four revised algebra measures in Districts B and C*. (Technical Report 10). Ames, IA: Iowa State University, Department of Curriculum and Instruction, Project AAIMS.

Foegen, A., Olson, J., & Perkmen, S. (2006). *An exploratory study of the use of two algebra progress monitoring measures to evaluate student growth (Technical report 11)*. Project AAIMS, Department of Curriculum and Instruction, Iowa State University, Ames, IA.

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

Fuchs, L. S., Hamlett, C. L., & Fuchs, D. (1998). *Monitoring basic skills progress: Basic math concepts and applications* [computer program manual]. Austin, TX: PRO-ED.

Hoffman, R. (n.d.). Assessing progressing monitoring measures in algebra. Unpublished manuscript, University of Houston, USA.

Hoover, H. D., Dunbar, S. B., Frisbie, D. A., Oberley, K. R., Bray, G. B., Naylor, R. J., et al. (2001). *The Iowa Tests of Basic Skills: Complete battery (Form A Levels 9-14).* Itaska, IL: Riverside Publishing.

Herscovics, N. (1989). *Cognitive obstacles encountered in the learning of algebra*. In Sigrid Wagner & Carolyn Kieran (Eds.), *Research issues in the learning and teaching of algebra* (pp. 60-86). Reston, VA: National Council of Teachers of Mathematics.

Item difficulty calculations. (n.d.). *In file help for Lertap 5*. Retrieved from http://lertap.curtin.edu.au/HTMLHelp/HTML/index.html?itemdifficultycalculations_.htm

Iowa End-of- Course assessments. (n.d.). *In Iowa testing programs*. Retrieved from http://www.education.uiowa.edu/itp/ieoc/Default.aspx

Kline, Paul (1986). *A handbook of test construction.* New York, NY: Methuen & Co.

Livingston, S. A. (2006). *Item Analysis*. In Downing & Haladyna (Eds.), *Handbook of test development* (pp. 421-441). Mahwah, NJ: Lawrence Erlbaum.

Matlock-Hetzel, S. (1997). *Basic concepts in Item and Test Analysis.* Paper presented at the annual meeting of the Southwest Educational Research Association, Austin, Texas. Retrieved 14 March 2009, from: http://www.eric.ed.gov/ERICDocs/data/ericdocs2sql/content_storage_01/0000019b/80/16/77/35.pdf

National Council of Teachers of Mathematics. (2006). *Curriculum focal points for prekindergarten through grade 8: A quest for coherence*. Reston, VA: NCTM.

Perkmen, S., Foegen, A., & Olson, J. (2006a). *Reliability, criterion validity, and sensitivity to growth: Extending work on two algebra progress monitoring measures.* (Technical Report 12). Ames, IA: Iowa State University, Department of Curriculum and Instruction, Project AAIMS.

Perkmen, S., Foegen, A., & Olson, J. (2006b). *A replication study of the reliability, criterion validity, and sensitivity to growth of two algebra progress monitoring measures.* (Technical Report 13). Ames, IA: Iowa State University, Department of Curriculum and Instruction, Project AAIMS.

Perkmen, S., Foegen, A., & Olson, J. (2006c). *Technical characteristics of two algebra progress monitoring measures: Reliability, criterion validity, and sensitivity to growth.* (Technical Report 14). Ames, IA: Iowa State University, Department of Curriculum and Instruction, Project AAIMS.

Salvia, J., & Yesseldyke, J. E. (1995). *Assessment* (6th ed.). Boston, MA: Houghton Mifflin Company.

Salvia, J., Ysseldyke, J.E. & Bolt, S. (2007).*Assessment in special and inclusive education*. (10th ed). Boston, MA: Houghton Mifflin Company.

Shapiro, E.S., Keller, M.A., Lutz, J.G., Santoro, L.E., & Hintze, J.M. (2006). Curriculum-based measures and performance on state assessment and standardized tests: Reading and math performance in Pennsylvania. *Journal of Psychoeducational Assessment*, *24*(1), 19-35.

Singamaneni, S., Foegen, A., & Olson, J. (2009). *Examining the long-term predictive validity of the Early Numeracy Indicators for predicting success on a high stakes mathematics test (Technical report 26)*. Minneapolis, MN: University of Minnesota, College of Education and Human Development, Research Institute on Progress Monitoring.

Stacey, K., & Chick, H. (2004). *Solving the problem with algebra*. In K. Stacey., H. Chick, & M. Kendal (Eds.), *The Future of Teaching and Learning of Algebra*. *The 12th ICMI Study* (pp. 1-20). Boston: Kluwer.

University of Washington, Seattle, Office of Educational Assessment. (August, 2005). *ScorePak®: Item analysis*. Retrieved 14 March 2009, from: http://www.washington.edu/oea/pdfs/resources/item_analysis.pdf