PROJECT AAIMS: ALGEBRA ASSESSMENT AND INSTRUCTION – MEETING STANDARDS

Reliability and Criterion Validity of
Five Algebra Measures in Districts B and C

Technical Report #7

Anne Foegen, Ph.D.
Jeannette Olson, M.S.
Serkan Perkmen, M.Ed.

Iowa State University

October 2005

Project AAIMS is funded by the U.S. Department of Education,
Office of Special Education Programs, Grant # H324C030060

PROJECT AAIMS: ALGEBRA ASSESSMENT AND INSTRUCTION – MEETING STANDARDS

Reliability and Criterion Validity of Five Algebra Measures in Districts B and C

Technical Report 7

Abstract

This technical report summarizes the results of a study in which we examined the technical adequacy of five potential measures for algebra progress monitoring. One hundred three students (14 of whom were receiving special education services) completed two forms of a *Basic Skills* measure, two forms of an *Algebra Foundations* measure, one form of a *Content Analysis-Constructed Response* measure, two forms of a *Translations* measure, and two forms of a *Content Analysis-Multiple Choice* measure administered over two data collection sessions. Each probe data collection session was repeated to investigate the test-retest reliability of the measures. In addition, we gathered data on criterion variables including grades, overall grade point average, teacher ratings of student proficiency, and scores on district-administered standardized tests, as well as a measure of algebra aptitude. We examined both test-retest and alternate form reliability for both single probe scores and aggregated scores (computed by averaging two individual scores). Criterion validity was examined by computing correlations between students' single and aggregated scores on the probes with their scores on other indicators of proficiency in algebra.

We found that four of the five measures produced effective distributions of student scores, with no signs of floor or ceiling effects. On the *Translations* probe, students produced nearly as many incorrect responses as they did correct responses, suggesting a high rate of guessing on that measure. The test-retest and alternate form reliability of single probes ranged from .4 to .9, with most coefficients in the .4 to .6 range. Aggregating scores from two probes produced slight increases in the reliability of the probes, with most correlations ranging from .5 to .7. For both single probes and aggregated scores, test-retest reliability coefficients exceeded those obtained for alternate form reliability. Neither the single nor the aggregated probes consistently produced reliability coefficients above the .80 level that represents a standard benchmark.

Criterion validity coefficients were also lower than those obtained in previous research (Foegen & Lind, 2004). Coefficients were generally in the low range (.2 to .4); the exception to this pattern was for the Iowa Algebra Aptitude Test, which was more strongly related to the algebra progress monitoring measures (coefficients in the .3 to .5 range). The Content Analysis-Constructed Response, the Algebra Foundations, and the Content Analysis-Multiple Choice measures produced the strongest relations with the criterion measures, with lower relations obtained for the *Basic Skills* and *Translations* measures.

Concerns were identified with difficulty of scoring the *Content Analysis-Constructed Response* probes efficiently and accurately, which will likely limit the viability of this measure in applied settings. Issues for future research are identified.

Full Report

Introduction

Algebra often functions in the role of a 'gatekeeper,' with proficiency in algebra having significant influence on individuals' access to higher education and professional career paths. If students with disabilities are to have access to these opportunities, it is critical that they develop proficiency in algebra. Robert Moses, a mathematics educator and civil rights advocate, sees algebra as the 'civil right' of the 21st century. He argues that algebra proficiency provides the same access to economic and social equity that the right to vote represented during the Civil Rights movement of the 1960s (Moses & Cobb, 2002). Project AAIMS (Algebra Assessment and Instruction—Meeting Standards) strives to improve student learning in algebra for all students, including those with and without disabilities. Project AAIMS has two primary objectives. First, we will examine the alignment between algebra curriculum, instruction, and assessment for students with and without disabilities. Second, we will develop and validate progress monitoring tools to support teachers' instructional decision making relative to student learning in algebra. In Technical Report 2 (Foegen & Lind, 2004), we reported the reliability and criterion validity of three measures developed as potential indicators of student proficiency in algebra. In Technical Report 6 (Foegen & Olson, 2005), we summarized the results of a study investigating two additional measures (for a total of five measures) in one of the three participating districts. In this report, we describe a study in two additional districts with all five of the potential measures of algebra proficiency.

Method

The study described in this report was conducted from October through December 2004 in Districts B and C. District B is located in a community of 26,000 people, where the high school currently serves 1,349 students. The majority of students are white (85%), and many are eligible for free and reduced lunch (23%). Thirteen percent of the students are of diverse backgrounds in terms of race, culture and ethnicity. A substantial proportion of the student population (15%, or about 202 students) is identified as eligible for special education services. District B uses block scheduling, so students complete a traditional course in approximately four and one half months. Each instructional period is 90 minutes in length, and the school day consists of four instructional periods.

District C is located in a predominantly rural area and serves approximately 17,700 residents in five small towns and a Native American Settlement community. The high school enrolls 488 students in grades 9 through 12. Twenty to 25% of the students are of diverse backgrounds in terms of race, culture and ethnicity. Approximately 44% of the school population is eligible for free and reduced lunch. Approximately 15% of the student population (73 students) has been identified as students eligible for special education services. Like District B, District C also uses block scheduling with a 90 minute period and four instructional periods in each school day.

Data for the study were gathered during an eight-week period from October through December 2004. During the first data collection session, students completed the algebra criterion measure. Two weeks later, the students began the four rounds of data collection for the algebra probes. All data collection activities involving students were completed during regular class time. Project AAIMS staff administered all measures.

Participants

Eighty-one students in District B and 22 students in District C participated in the study. Written parental/guardian consent and written student assent were obtained for all of these students using procedures approved by Iowa State University's Human Subjects Review Committee. Descriptions of the participating students from each district are provided in Tables 1 and 2.

Table 1. Demographic Characteristics of Student Participants by Grade Level for District B

	Total	Grade 9	Grade 10	Grade 11	Grade 12
N	81	35	34	10	2
Gender					
Male	33	15	14	3	1
Female	48	20	20	7	1
Ethnicity					
White	70	30	29	9	2
Black	9	3	5	1	0
Hispanic	1	1	0	0	0
Native Am.	0	0	0	0	0
Lunch					
Free/Red	34	15	13	5	1
Disability					
IEP	14	1	8	5	0

Table 2. Demographic Characteristics of Student Participants by Grade Level for District C

	Total	Grade 9	Grade 10	Grade 11	Grade 12
N	22	16	3	2	1
Gender					
Male	10	10	0	0	0
Female	12	6	3	2	1
Ethnicity					
White	17	13	2	2	0
Black	0	0	0	0	0
Hispanic	2	0	1	0	1
Native Am.	3	3	0	0	0
Lunch					
Free/Red	8	3	2	2	1
Disability					
IEP	0	0	0	0	0

As the data in Tables 1 and 2 indicate, many of the participants (an average of 85%) were white and an average of 50% were in ninth grade, the traditional grade in which students in these districts complete algebra. Forty-two and 36% participated in federal free and reduced lunch programs in Districts B and C, respectively and 17% of the participating students in District B were students with disabilities who were receiving special education services. In District B, 56 students were enrolled in Algebra 1A and 25 were in Algebra 1B   Algebra 1A/1B is an option available in District B in which students complete half the content of a traditional Algebra 1 course in a single course. In District C, all the students were enrolled in Algebra 1. Due to the small number of students in District C participating in the study, data from students in the two schools were combined for statistical analyses purposes.

Additional Information on Students with Disabilities. Because the applicability of the algebra probes to students with disabilities is an important part of Project AAIMS, additional information about the 14 students with disabilities in District B participating in the project is provided in Table 3.

Students with disabilities earned mean grades in algebra of 1.86 [C-] (range 0.00 [F] to 4.00 [A]). In Districts B and C, the Iowa Tests of Educational Development are used as a district-wide assessment. On average, students with disabilities obtained national percentile rank score of 31 and 34 in Concepts/Problem Solving, and Computation, respectively. They demonstrated greater deficits in reading, with a mean percentile rank of 26 for the Reading Total score.

Table 3. Descriptive Information on the Programs of Students with Disabilities in District B

Characteristic	Quantification
Disability category	12 Entitled Individual (EI) 2 no additional information available
% time in general education	All reporting students spend 100% of their instructional time in general education
# of students with math goals	7
# of students receiving math instruction in general education classes	14
# of students with goal code F2C: *Comprehension*	7
# of students with goal code F3A: *Applied math*	1
# of students with goal code F3C*:* *Computation*	5
# of students with goal code F4*: Demonstrates competence in basic written language skills*	1
# of students with goal code F4M: *Mechanics of writing: punctuation, grammar, spelling*	8

Measures

Two groups of measures were used in this study. The first group consists of the curriculum-based measures of algebra performance developed by the Project AAIMS research team. The second group consists of the measures that served as criterion indicators of students' proficiency in algebra. Each group of measures is described below.

Algebra Progress Monitoring Measures. Five algebra measures were examined in this study; copies of each are provided in the Appendix. The following paragraphs summarize the characteristics of each of the five types of algebra measures.

Probe A: Basic Skills Measure

Probe A was designed to assess the 'tool skills' that students need to be proficient in algebra. Just as elementary students' proficiency with basic facts is associated with their ease in solving more complex problems, we hypothesized that there are some basic skills in algebra that serve as indicators of overall proficiency. In our discussions with algebra teachers, they frequently commented that many students had difficulty with integers and with applying the distributive property. The items included in the *Basic Skills* measure addressed solving simple equations, applying the distributive property, working with integers, and combining like terms. The *Basic Skills* probe included many skills one would assume that students proficient in algebra would be able to complete with reasonable levels of automaticity. Each *Basic Skills* probe consisted of 60 items; each item was scored as one point if it was answered correctly.

Probe B: Algebra Foundations Measure

The second algebra progress monitoring probe was the *Algebra Foundations* measure. This assessment was designed to reflect five core concepts and skills that we derived from our reading of the literature and our conversations with colleagues in mathematics education. The five foundation areas included (1) writing and evaluating variables and expressions; (2) manipulating expressions; (3) graphing expressions and linear equations; (4) solving one-step equations and simplifying expressions; and (5) identifying and extending patterns and functions. Our intent with this measure was to assess the extent to which students are proficient in solving problems that address these foundations of early algebra. It is important to note that in this measure, many of the items represent concepts and skills that would be learned as part of pre-algebra or very early instruction in an Algebra I course, if not earlier. We recognized that proficiency on this measure is not equivalent to having mastered all the concepts taught in Algebra I, but we hoped to determine whether the scores for this measure might serve as an indicator of more general proficiency in algebra. The *Algebra Foundations* probe consisted of 42 items; eight of these items required two responses, so 50 total points were possible on this probe.

Probe C: Content Analysis-Constructed Response Measure

The third measure was the *Content Analysis-Constructed Response* probe. We developed two forms of this measure by analyzing the content included in the algebra instructional materials. Since all three districts participating in Project AAIMS were using the same textbook series, we wanted to investigate a measure that was directly derived from their common instructional materials. We developed the items by sampling from the chapter tests and reviews. We sought to identify items that represented core concepts/problem types in each chapter. Based on teacher feedback, we sampled chapters in the middle portion of the text at a higher rate than the chapters at the beginning (review) and end (advanced concepts/skills) sections of the text. We anticipated that this probe might provide a more direct reflection of the extent to which students have mastered algebra content than would the other probes, which represent more general indicators of algebra proficiency.

The *Content Analysis-Constructed Response* probes consisted of 16 items. Because all of the textbook chapters addressed more than one core concept, different forms of this probe tested different big ideas for each chapter; consequently, the number of points possible on the probes varied from 53 for Form C-1 to 59 for Form C-2. The number of points possible for the individual problems ranged from one to six. This probe required more complex problem solving and generated fewer responses; therefore, we needed a means of increasing the sensitivity of the scores. We asked the algebra teachers to solve each of the problems, showing the steps and the final solution in the same way that they would expect students to show their work. We then assigned a point to each step of the solution. Students were awarded points corresponding to any of the steps they complete correctly. In the directions for this probe, we encouraged students to show their work to obtain 'partial credit' even if they weren't able to solve the entire problem.

Probe D: Translations Measure

The fourth probe, which we referred to as the *Translations* probe, was designed to assess the students' proficiency in recognizing translations between multiple representations of the relationships between two sets of numbers. In creating this probe, we drew from curriculum materials for teaching algebra concepts at the middle school level created as part of the *Connected Mathematics* project (Lappan, Fey, Fitzgerald, Friel, Phillips, 2004). In this curriculum program, students explore the connections between numerical relationships in multiple formats. For example, they might examine how changing elements of an equation (i.e., changing $y = 2x$ to $y = 2x + 3$) influences the graphic representation of the equation. Likewise, they examine relationships between data tables, graphs, and equations. Contextualized problems representing real life situations are also used as a basis for exploring algebraic relationships. In our *Translations* probe, we assessed whether students could recognize the same relationship between two sets of numbers presented in four different formats. At the top of the first page, students were given four 'base' graphs (on the second page, equations were used as the stimulus and on the third, data tables). Below these four prompts (labeled A through D), students were presented with rows of alternative representations of the same relationships. One row contained equations, another data tables, and a third, story scenarios. The students' task was to identify matches between the four prompts at the top of the page and the same relationships represented in another format in each of the following three rows.

The *Translations* probe was created in response to feedback from the Project AAIMS Advisory Committee during a review of the initial three algebra probes. The Advisory Committee noted that the initial three probes focused heavily on algebraic manipulations and procedures, and urged the AAIMS research staff to pursue the development of a task that allowed students to demonstrate conceptual understandings of algebraic topics without requiring procedural accuracy with manipulations of algebraic symbols. In order to fit with the design constraints for progress monitoring tasks (i.e., brief, easy to administer and score), we selected a multiple choice format for the task. We created two parallel forms of the *Translations* probes. Each probe consisted of 42 items; we scored the probes by counting the number of correct and incorrect responses. Because of the multiple choice format, we were concerned that scores might be artificially inflated by guessing. Previous work by Foegen (2000) has demonstrated that applying a correction formula for guessing increases the reliability and criterion validity of the scores.

Probe E: Content Analysis-Multiple Choice Measure

The fifth algebra progress monitoring measure that we developed was the *Content Analysis-Multiple Choice* measure. This measure was a variation of the *Content Analysis-Constructed Response* (C) probe examined in the study reported in Technical Report 2 (Foegen & Lind, 2004). For this measure, we revised the original probe by creating four multiple-choice alternatives for each problem. Our rationale for going to a multiple choice option was that this format would improve scoring efficiency (and potentially interscorer agreement), that it might reduce the difficulty of the task (on the open ended version of the probe, we obtained significant floor effects, even when the probe was administered at the end of a year of instruction), and that the multiple choice format was one that students needed to be proficient with for district-administered assessments. Students had seven minutes to complete the probe; they were encouraged to show their work in order to earn partial credit even if they were not able to completely solve a problem. In addition, students were advised NOT to make wild guesses, as these would result in deductions from their total scores.

Scoring for the *Content Analysis-Multiple Choice* probes was done by comparing student responses to a rubric-based key created by the research staff. Each of the 16 problems was worth up to three points. Students earned full credit (three points) by circling the correct answer from among the four alternatives. If students circled an incorrect response and did not show any work, their answer was considered a 'guess' and counted as part of the final score assigned to each probe. In cases where students showed work, the scorer compared the student's work to the rubric-based key, and determined whether the student had earned 0, 1, or 2 points of partial credit. A student's final score on the probe consisted of the number of points earned across all 16 problems. The number of guesses was also recorded and entered in the data files.

Copies of each of the five types of probes used in the study are included in the Appendix.

Criterion Measures. In order to evaluate the criterion validity of the algebra progress monitoring measures, we gathered data on a variety of other indicators of students' proficiency in algebra. Some of these measures were based on students' performance in class (and in school more generally) and their teachers' evaluation of their proficiency. Other measures reflected students' performance on standardized assessment instruments.

The classroom-based measures included grade-based measures and teacher ratings. Each student's *algebra grade*, the grade s/he earned in algebra during the fall semester of the 2004-05 school year, was recorded using a four-point scale (i.e., A = 4.0, B = 3.0). *Grade point averages* for fall were available for students in District C; these were reported using the same four-pint scale. We also wanted to include the teachers' evaluations of students' proficiency in algebra. To accomplish this, we asked each teacher to complete a *teacher rating* form for all the students to whom s/he taught algebra. Student names were alphabetized across classes to minimize any biases that might be associated with particular class sections. Teachers used a 5-point Likert scale (1=low proficiency, 5= high proficiency) to rate each student's proficiency in algebra in comparison to same-grade peers. A copy of the teacher rating form is included in the Appendix.

Student performance on standardized, norm-referenced assessments was evaluated using school records and with an algebra instrument administered as part of the project. In Districts B and C, students complete the *Iowa Tests of Educational Development* (ITED) as a district-wide assessment. District records were used to access students' scores on these instruments; national percentile ranks were used for the analyses. Students in District B completed the ITED in November 2004, while students in District C completed the ITED in March 2005. We recorded the Concepts/Problems score (which was identical to the Math Total score), the Computation score, and the Reading Total score.

Because the district-administered measure did not provide a direct assessment of algebra, so we also administered the *Iowa Algebra Aptitude Test* (IAAT). This norm-referenced instrument is typically used to evaluate the potential of 7^{th} grade students for successful study of algebra in 8^{th} grade. Although we recognized the limitations of using this aptitude measure, we were unable to identify a norm-referenced test of algebra achievement. We had some concerns that there might be ceiling effects when using this measure, but these concerns proved to be unwarranted.

Procedures

The algebra probes were administered during a portion of each class period. Because Districts B and C use block scheduling, each period was approximately 90 minutes in length. To avoid potential problems with fatigue, we separated the five different types of algebra probes into two groups. One group consisted of two parallel forms of the *Basic Skills* probe, two forms of the *Translations* probe, and one form of the *Content Analysis-Constructed Response* probe. The second group included two parallel forms of the *Algebra Foundations* probe and two parallel forms of the *Content Analysis-Multiple Choice* probe. The order in which the probes were administered was counterbalanced across classes, as was the order of each of the parallel forms. The one exception to this procedure was the *Content Analysis-Constructed Response* probe, which was always administered at the end of the first group of probes. Students completed the tasks in the same order both weeks. A copy of the standardized directions used for each administration session is provided in the Appendix. Table 4 depicts the order in which the probes were administered during each of the two testing sessions. The two probes indicated for the *Content Analysis-Constructed Response* probes reflect the form the students completed the first week and the second week, respectively.

Table 4. Administration Schedule for Probe Forms by Period

Session	E-1 (1A)	E-3 (1B)	E-4 (1B)	K-2 (1A)	K-3 (1A)	K-4 (1A)	B-1 1	B-4 1	C-1 1
1 and 2	A1	D1	A2	B1	E1	B2	D2	A1	E2
	A2	D2	A1	B2	E2	B1	D1	A2	E1
	D1	A1	D2	E1	B1	E2	A2	D2	B2
	D2	A2	D1	E2	B2	E1	A1	D1	B1
	C1/C2	C2/C2	C1/C1				C2/C1	C1/C2	
3 and 4	B1	E1	B2	A1	D1	A2	E2	B1	D2
	B2	E2	B1	A2	D2	A1	E1	B2	D1
	E1	B1	E2	D1	A1	D2	B2	E2	A2
	E2	B2	E1	D2	A2	D1	B1	E1	A1
				C1/C2	C2/C2	C1/C1			C1/C1

A1, A2 = *Basic Skills* probes 1 and 2
B1, B2 = *Algebra Foundations* probes 1 and 2
C1, C2 = *Content Analysis-Constructed Response* probes 1 and 2
D1, D2 = *Translations* probes 1 and 2
E1, E2 = *Content Analysis-Multiple Choice* probes 1 and 2

Scoring

The scoring for the *Basic Skills* and *Algebra Concepts* probes was completed by counting the number of problems completed correctly. For these two probe types, we continued the investigation of alternative durations that we began in the study reported in Technical Report 2. As students completed the *Basic Skills* probe, they were instructed to move down the columns or across the rows. After two minutes had elapsed, the research team member administering the tasks said “SLASH” and students made a dark slash mark on the problem on which they were working. This process was repeated at the end of three minutes and students were instructed to stop working on the probe at the end of four minutes. A similar procedure was used with the *Algebra Foundations* probe, with durations of four, five, and six minutes to work on the probe.

For each probe, the scoring process included determining the number of problems completed in each of the three durations.

The scoring for the *Content Analysis-Constructed Response* probe involved comparing each student's response to the key generated from the teachers' solutions of the problems. Students were awarded one point for each step of the problem that they completed correctly. Students who had a correct solution to a problem were awarded full credit, regardless of the number of sub-steps they showed in their work. The total score for this probe was the number of points earned across all 18 problems.

The scoring for the remaining two probes, *Translations* and *Content Analysis-Multiple Choice*, was identical to the procedures described in Technical Report 6. We recorded the number of correct and incorrect responses provided by each student for each probe, then examined three alternative scoring procedures in our analyses. The first scoring method involved using the total points earned on the probe (we refer to this value as the 'Correct' score in subsequent analyses). The second method (which we've labeled C – I in subsequent analyses) involved subtracting the number of Incorrect problems from each student's total Correct points. The third method (labeled 'C - 1/3' in subsequent analyses) involved subtracting one third of the number of incorrect problems from the total points earned on each probe. This procedure to correct for guessing has been used in previous research involving multiple choice mathematics probes and was found to be effective in increasing the reliability and validity of the scores (Foegen, 2000). In circumstances where the scoring procedure produced a negative value, the student's score was set to 0.

Results

Scoring Reliability

Scoring accuracy was evaluated by re-scoring approximately one-third of the probes. For each probe, an answer-by-answer comparison was conducted and an interscorer reliability estimate was calculated by dividing the number of agreements by the total number of answers scored. These individual probe agreement percentages were then averaged across all the selected probes of a common type to determine an overall average.

We selected the probes to be re-scored by sampling from the nine class periods across the four administration periods. Each form of the probes was rescored for at least 6 of the 18 administrations (33%). The one exception was the *Content Analysis-Constructed Response* probe, which was rescored for 4 of the 9 periods in which it was administered. The number of student papers rescored and the average agreement for each form of the probe are reported in Table 5 below.

Table 5. Interscorer Agreement Rates and Student Papers Rescored

Probe	# Papers Rescored	Range of Agreement	Mean % Agreement
Basic Skills	193	81 – 100%	98.8
Algebra Foundations	177	81 – 100%	97.7
Content Analysis-Constructed Response	128	63 – 100%	91.9
Translations	247	81 – 100%	99.3
Content Analysis-Multiple Choice	184	73 – 100%	97.6

As in our earlier study, the *Content Analysis-Constructed Response* measure proved to be the most difficult to score consistently across multiple scorers. In addition, we found that this measure required substantially more time per probe to score, which will limit its viability for secondary general education teachers, many of whom interact with large numbers (sometimes more than 100) students per day. All of the other measures could be scored with acceptable levels of agreement between scorers.

Descriptive Data on Score Ranges and Distributions

Table 6 lists the ranges, means, and standard deviations for each of the probes. For the *Basic Skills* and *Algebra Foundations* probes, data are reported for each of the three alternative durations we investigated. On the *Translations* probe, the Correct score represents the number of correct matches, while the Incorrect score represents the number of incorrect responses. The total possible for the *Translations* probe was 42 points. On the *Content Analysis-Multiple Choice* probes, the Correct score represents the number of points earned on the probe (each of the 16 problems was worth up to 3 points) and the Incorrect score represents the number of incorrect responses.

Table 6. Descriptive Data for Algebra Probes Across Administration Sessions – Raw Scores

Measure	Session/ Week	N	Score Type	Range	Mean Score	Standard Deviation
Basic Skills Form 1	1	91	2 min.	0 - 22	9.48	3.91
		91	3 min.	0 - 29	13.69	5.50
		92	4 min.	0 - 38	17.59	7.17
	2	92	2 min.	0 - 23	11.14	4.66
		92	3 min.	0 - 38	15.70	6.65
		92	4 min.	0 - 42	19.77	8.05
Basic Skills Form 2	1	91	2 min.	0 - 17	8.69	3.33
		91	3 min.	0 - 26	12.22	4.84
		91	4 min.	1 - 35	15.92	6.43
	2	93	2 min.	0- 21	10.00	3.98
		93	3 min.	0 - 33	14.10	5.75
		93	4 min.	0- 44	18.32	7.73
Algebra Foundations Form 1	1	89	4 min.	2 – 28	13.90	4.43
		88	5 min.	2 – 32	16.23	4.97
		89	6 min.	2 – 38	18.38	5.78
	2	69	4 min.	5 – 28	18.00	4.83
		69	5 min.	6 – 35	20.43	6.04
		69	6 min.	7 – 40	23.70	6.45
Algebra Foundations Form 2	1	77	4 min.	0 – 29	13.00	5.79
		91	5 min.	0 – 32	16.41	6.18
		92	6 min.	0 – 37	18.76	6.73
	2	70	4 min.	3 – 32	17.29	5.78
		70	5 min.	4 – 38	20.53	6.58
		70	6 min.	5 – 43	23.54	7.67
Content Analysis-Constructed Response – Form 1 (C1)	1	54	Points earned	5 – 34	16.96	5.27
	2	51		10 – 34	17.02	4.76
Content Analysis-Constructed Response – Form 2 (C2)	1	38	Points earned	6 – 30	16.11	5.74
	2	33		2 – 33	15.36	7.55

Table 6. Descriptive Data for Algebra Probes Across Administration Sessions – Raw Scores (continued)

Measure	Session/ Week	N	Score Type	Range	Mean Score	Standard Deviation
Translations Form 1	1	92	Correct	1 – 25	10.82	4.84
		92	Incorrect	0 – 34	8.97	8.31
	2	91	Correct	1 – 26	12.78	5.79
		91	Incorrect	0 – 37	10.75	9.19
Translations Form 2	1	93	Correct	2 – 27	12.00	5.11
		93	Incorrect	0 – 35	8.61	8.47
	2	91	Correct	1 – 34	12.89	6.45
		91	Incorrect	0 – 35	10.15	9.43
Content Analysis-Multiple Choice Form 1	1	90	Correct	5 – 36	20.27	6.18
		90	Incorrect	0 – 11	4.77	2.79
	2	85	Correct	8 – 39	22.27	6.27
		85	Incorrect	0 – 12	4.36	3.04
Content Analysis-Multiple Choice Form 2	1	92	Correct	3 – 31	18.64	6.06
		92	Incorrect	0 – 13	3.17	3.15
	2	84	Correct	6 – 34	20.26	6.54
		84	Incorrect	0 – 13	3.00	3.09

Results for the *Basic Skills* and *Algebra Foundations* probes revealed increasing mean scores and standard deviations with each one-minute increase in duration. This finding suggests that longer durations will more effectively spread out students for the purpose of gauging their performance in algebra. Although some students obtained scores of 0 on both types of probes, this was an infrequent occurrence and we did not interpret this as evidence of a floor effect. No problems with ceiling effects were observed. Likewise, the *Content Analysis-Constructed Response* probe did not show any evidence of floor or ceiling effects.

The results for the *Translations* probes showed a pattern similar to the results reported in Technical Report 6 (Foegen & Olson, 2005). The mean number of incorrect problems was nearly as large as the number of correct problems. We interpreted these results as evidence of a high rate of student guessing on this measure. As in District A in the earlier study, the conceptual understanding of algebra necessary to complete the *Translations* probe successfully was not emphasized in the instructional materials or in the teachers' obstruction we observed. We also noted more signs of student frustration during the administration of the *Translations* probe than for any of the other four measures.

The *Content Analysis-Multiple Choice* results did not demonstrate any problems with floor or ceiling effects. The mean number of points earned (Correct) exceeded the number of problems incorrect by four to five times. In addition, the standard deviations for the Correct scores were about double those for the Incorrect scores.

Reliability of Individual Probe Scores

The reliability of individual probes was evaluated by examining alternate form reliability (the Pearson product moment correlation between the two forms of each type of probe given during the same data collection session) and test-retest reliability (the Pearson correlation between the same form of a probe given across the two data collection sessions). In Table 7, we summarize the reliability data for individual probe scores for each of the five types of algebra probes.

Table 7: Reliability results for single probes

Probe Type		Alternate Forms				Test-Retest		
Basic Skills		2 min.	3 min.	4 min.		2 min.	3 min.	4 min.
Session 1		.59	.57	.65	Form 1	.65	.66	.67
Session 2		.64	.78	.81	Form 2	.68	.67	.72
Algebra Foundations		4 min.	5 min.	6 min.		4 min.	5 min.	6 min.
Session 1		.41	.48	.53	Form 1	.72	.72	.76
Session 2		.58	.53	.66	Form 2	.64	.70	.71
Content Analysis-Constructed Response[a]			Corr.				Corr.	
Session 1			.70		Form 1		.44	
Session 2			NS		Form 2		.92	
Translations		Corr	C – I	C -1/3		Corr	C – I	C -1/3
Session 1		.47	.44	.47	Form 1	.43	.49	.50
Session 2		.66	.69	.68	Form 2	.58	.61	.62
Content Analysis-Multiple Choice		Corr	C – I	C -1/3		Corr	C – I	C -1/3
Session 1		.47	.47	.45	Form 1	.56	.54	.54
Session 2		.61	.57	.59	Form 2	.56	.56	.57

Note: All correlations significant at $p < .05$.; NS = non-significant

[a]for this probe only, alternate form reliability was computed by correlating students' scores on one form during session one and a different form during session two.

For the *Basic Skills* probes, alternate form reliability coefficients ranged from .5 to .8, and test-retest reliability coefficients ranged from .6 to .7. In general, increases in the duration of the measure produced reliability coefficients that were similar or larger than those for the shorter duration; in both cases and across both sessions of data collection, the four-minute duration produced the largest coefficients.

For the *Algebra Foundations* probes, test-retest reliability coefficients were larger than the coefficients for alternate form reliability (.6 to .7 and .4 to .6, respectively). As with the *Basic Skills* probes, the longest duration produced the highest reliability coefficients. It must be noted that the reliability estimates for single probes for both *Basic Skills* and *Algebra Foundations* do not consistently meet the standard expectation of .80 or higher.

Reliability estimates for the *Content Analysis-Constructed Response* probes were highly variable. Because students completed only one form of this measure in each data collection session, it was not possible to compute alternate form reliability coefficients in the same manner as for the other measures. Instead, the data reported in Table 7 for alternate form reliability represent the correlation coefficient between students who completed Form 1 in the first session and Form 2 in the second session (Session 1 in the table) and Form 2 in the first session and Form 1 in the second session (Session 2 in the table). One of these coefficients was not statistically significant, the other was .70. For test-retest reliability, the estimate for Form 1 was .44, while the estimate for Form 2 was .92. These highly variable results raise questions about the reliability of this measure.

Because the *Translations* and *Content Analysis-Multiple Choice* probes both used a multiple-choice format, we were concerned about the potential that guessing might artificially inflate students' scores. We compared the effects of three different scoring procedures on the reliability of students' scores on the probes. The first scoring method involved using the total points earned on the probe (i.e., the values listed in Table 5 as 'Correct'). Findings for this scoring method are listed under the column titled Correct in Table 7. The second method (listed in the column titled C – I in Table 7) involved subtracting the number of incorrect problems (the 'Incorrect Value' in Table 6) from each student's total Correct points. The third method (labeled '1/3' in Table 7) involved subtracting one third of the number of incorrect problems from the total points earned on each probe. This procedure to correct for guessing has been used in previous research involving multiple-choice mathematics probes and was found to be effective in increasing the reliability and validity of the scores (Foegen, 2000). In circumstances where the scoring procedure produced a negative value, the student's score was set to 0.

Reliability coefficients for the *Translations* and *Content Analysis-Multiple Choice* probes were in the .4 to .6 range. In general, students' scores were more consistent across forms during the second data collection session and (for *Translations* only), for Form 1 more than for Form 2 (though readers should recall the order of the forms was counterbalanced across classes). Mixed results were obtained regarding the use of scoring procedures to correct for guessing. For both measures, the reliability estimates for single probes are well below expected standards.

<u>Reliability of Aggregated Probe Scores</u>

Because students completed two forms of most probes during each data collection session, it was also possible to examine the effects of aggregating scores from two probes on the resulting reliability levels. Previous research in other areas of mathematics (Foegen, 2000; Fuchs, Deno, & Marston, 1983) has determined that for some types of mathematics skills and concepts, multiple probes need to be aggregated to obtain reliable scores for individual students. Table 8 presents the results for the aggregated scores on probes. The alternate form coefficients were computed by averaging scores from the two administrations of Form 1, then the average of

Table 8. Reliability for Aggregated Probes

Probe	Alternate Form Reliability	Test-Retest Reliability
Basic Skills		
2 min.	.67	.75
3 min.	.73	.73
4 min.	.78	.75
Algebra Foundations		
4 min.	.51	.69
5 min.	.55	.77
6 min.	.61	.77
Translations		
Correct	.68	.64
C - I	.69	.69
C - 1/3	.69	.69
Content Analysis-Multiple Choice		
Correct	.58	.61
C - I	.58	.60
C - 1/3	.57	.61

Note: All correlations significant at $p < .05$.; NS = non-significant

scores from the two forms of each probe administered on the first data collection day, and then the two administrations for Form 2. The test-retest coefficients were computed by averaging correlating these scores with the averaged scores for the same probes from the second data collection day.

For the *Basic Skills* and *Algebra Foundations* probes, aggregating by averaging students' scores on two forms of a probe generally produced improvements for test-retest reliability. The same pattern was not obtained for alternate form reliability, where some results based on aggregation increased, while others were similar to the results obtained for single probes. For the *Translations* and the *Content Analysis-Multiple Choice* measures, substantial improvements were obtained for test-retest reliability. Again, only mixed results were obtained for test-retest reliability. For these two measures, aggregating scores increased the reliability coefficients into the .5 to .6 range, but this is not a sufficient improvement relative to expected standards. Future research should explore additional revisions to the design of the measures and the administration procedures to address these limitations in their technical adequacy.

Criterion Validity for Single Probes

The criterion validity of the measures was examined by correlating scores on the probes with the criterion measures that served as additional indicators of students' proficiency in algebra. The indicators we used included students' grades in algebra and (for District C only), fall semester GPA; teachers' evaluations of student proficiency; scores from standardized tests in mathematics administered by the district; and scores obtained from a norm-referenced test of algebra aptitude, the Iowa Algebra Aptitude Test (IAAT). In the following section, the correlation coefficients between scores on the algebra measures and each of these criterion variables are presented and discussed. Correlation coefficients for the grade-based criterion measures and for teacher ratings are presented in Table 9, with results included for each of the three timing options for the *Basic Skills* and *Algebra Foundations* probes and each of the three scoring methods used for the *Translations* and *Content Analysis-Multiple Choice* probes. Because four correlation coefficients were produced in the analyses (scores from each of two forms of probe were available for each of the two administration days) for most of the probes, mean correlations are reported. The range of obtained correlations is included in parentheses. If at least two of the four correlations were statistically significant, the mean correlation is reported. Table 10 includes the correlations between the ITED subtests and the algebra measures, while we report the correlations between the algebra measures and the IAAT subtests and total score in Table 11.

Table 9. Correlations between algebra probes and grade-based measures and teacher ratings

		Overall GPA	Grade in Algebra		Teacher Rating	
Basic Skills	2 min.	NS	.27	(2 NS, .25 - .28)	.30	(.21 - .36)
	3 min.	NS	.23	(1 NS, .22 - .26)	.31	(.23 - .34)
	4 min.	NS	.23	(2 NS, .22 - .23)	.31	(.25 - .37)
Algebra Foundations	4 min.	NS	.38	(2 NS, .32 - .43)	NS	
	5 min.	NS	.36	(2 NS, .28 - .44)	NS	
	6 min.	NS	.32	(1 NS, .22 - .46)	NS	(3 NS; .22)
Content Analysis – Constructed Response		NS	.40	(2 NS, .39 - .41)	.37	(1 NS, .27 - .54)
Translations	Correct	NS	NS	(3 NS; .35)	NS	(3 NS; .23)
	C – I	NS	.36	(.34 - .48)	.27	(.21 - .34)
	1/3	NS	.30	(1 NS, .23 - .47)	NS	(3 NS; .29)
Content Analysis-Multiple Choice	Correct	NS	NS	(3 NS; .32)	.28	(2 NS, .27 - .29)
	C – I	NS	.34	(2 NS, .30 - .37)	.31	(2 NS, .29 - .32)
	1/3	NS	.30	(2 NS, .24 - .35)	.29	(2 NS, .29 - .29)

Correlations between the five algebra measures and the grade-based measures and teacher ratings were in the low range (.2 to .4), with many not achieving statistical significance. The strongest relations were identified with the *Content Analysis-Constructed Response* measure. In previous studies, overall GPA has demonstrated stronger relations with the algebra measures than algebra grades alone. Our results here found that overall GPA was not related to students' performance on any of the measures. In addition, the correlations involving teacher ratings were lower in this set of results than for the results obtained in the studies reported in Technical Report 2 (Foegen & Lind, 2004) or in Technical Report 6 (Foegen & Olson, 2005).

Table 10. Correlations between algebra measures and ITED scores

		ITED Scores		
		Con/Prob	Comp	Reading Total
Basic Skills	2 min.	NS (3 NS; .26)	.28 (1 NS, .24 - .32)	NS
	3 min.	NS (3 NS; .25)	.26 (1 NS, .24 - .29)	NS
	4 min.	NS (3 NS; .26)	.26 (.23 - .29)	NS
Algebra Foundations	4 min.	.35 (2 NS, .27 - .43)	NS	.33 (1 NS, .25 - .38)
	5 min.	.33 (1 NS .22 - .45)	.25 (1 NS, .24 - .26)	.32 (.25 - .40)
	6 min.	.31 (1 NS, .23 - .37)	.28 (2 NS, .27 - .29)	.32 (.28 - .35)
Content Analysis – Constructed Response		.44 (2 NS, .35 - .53)	.50 (2 NS, .46 - .54)	.40 (2 NS, .39 - .41)
Translations	Correct	NS	NS	NS
	C – I	.25 (2 NS, .22 - .27)	NS (3 NS; .28)	NS
	1/3	NS	NS	NS
Content Analysis-Multiple Choice	Correct	.33 (2 NS, .32 - .34)	.25 (2 NS, .24 - .25)	NS (3 NS; .23)
	C – I	.32 (2 NS, .30 - .33)	NS (3 NS; .25)	NS (3 NS; .22)
	1/3	.33 (2 NS, .31 - .34)	.23 (2 NS, .22 - .24)	NS (3 NS; .23)

Correlations between students' scores on the probes and their scores on the Iowa Test of Educational Development (Table 10) were often non-significant, and those correlation coefficients that were significant were extremely low (.2 to .3 range). The one exception to this pattern was the *Content Analysis-Constructed Response* probe, which produced correlations in the low to moderate range (.4 to .5)

Table 11 reports the results of correlations between students' probe scores and their scores on the four subtests and the total scale score of the Iowa Algebra Aptitude Test. These correlations reflected stronger relations than those obtained with the ITED. This result is not surprising, as the content of the ITED is more focused on general mathematics proficiency (computation and problem solving) than on algebraic thinking. Again, the strongest relations were obtained with the *Content Analysis-Constructed Response* measure, with the *Algebra Foundations* measure also demonstrating relations in the low to moderate range (.4 to .5) with the total scale score on the IAAT. The next strongest measure was the *Content Analysis-Multiple Choice*, followed by the *Basic Skills* measure. The *Translations* measure had the weakest relation with student performance on the IAAT, further substantiating our hypothesis that students responded to this measure largely by guessing or providing random responses.

Table 11. Correlations between algebra progress measures and the IAAT subtests and total test scores

	Total	Part A	Part B	Part C	Part D
Basic Skills					
2 min.	.39 (.35 - .43)	NS (3 NS; .21)	.32 (.27 - .37)	.30 (.27 - .37)	.30 (1 NS, .22 - .35)
3 min.	.36 (.31 - .44)	NS	.30 (.22 - .37)	.28 (.23 - .32)	.34 (2 NS, .29 - .39)
4 min.	.38 (.28 – 47)	NS	.34 (1 NS, .29 - .37)	.30 (.28 - .34)	.32 (1 NS, .21 41)
Algebra Foundations					
4 min.	.46 (.33 - .57)	.31 (2 NS, .27 - .34)	.38 (.31 - .45)	.37 (.31 - .44)	.38 (2 NS, .35 - .40)
5 min.	.50 (.43 - .59)	.30 (1 NS, .27 - .36)	.41 (.37 - .45)	.39 (.32 - .46)	.36 (1 NS, .29 - .42)
6 min.	.51 (.45 - .58)	.31 (1 NS, .27 - .37)	.40 (.34 - .48)	.43 (.41 - .45)	.34 (1 NS, .32 - .36)
Content Analysis – Constructed Response					
	.56 (1 NS, .48 - .68)	.45 (2 NS, .43 -.47)	.45 (1 NS, .38 - .51)	.41 (1 NS, .34 - .50)	.40 (1 NS, .28 - .51)
Translations					
Correct	NS (3 NS; .34)	NS (3 NS; .32)	NS	.27 (2 NS, .23 - .31)	NS
C – I	.34 (.24 - .46)	.29 (1 NS, .21 - .39)	.26 (1 NS, .22 - .28)	.30 (2 NS, .21 - .38)	.27 (1 NS, .26 - .29)
1/3	.30 (.23 - .37)	.32 (2 NS; .27 - .37)	.23 (1 NS; .22 - .25)	.29 (2 NS; .25 - .32)	NS
Content Analysis-Multiple Choice					
Correct	.43 (.33 - .53)	.28 (1 NS, .25 - .30)	.32 (.23 - .39)	.31 (1 NS, .29 - .34)	.35 (1NS, .33 - .38)
C – I	.43 (.33 - .52)	.28 (.25 - .30)	.33 (.22 - .38)	.33 (1 NS, .31 - .35)	.34 (1 NS; .32 - .36)
1/3	.44 (.34 - .54)	.28 (.23 - .31)	.33 (.23 - .40)	.32 (1 NS, .31 - .35)	.35 (1 NS, .34 - .37)

Criterion Validity for Aggregated Probe Scores

In Tables 12 to 14, we report the criterion validity coefficients using the same set of criterion variables and the aggregated scores for each of the probes. To aggregate, we averaged the two scores of a probe type that were administered on the same day. This produced two scores for the Basic Skills and Algebra Concepts probes (Day 1 aggregate, Day 2 aggregate). We also aggregated scores by computing the average score on the same form administered across two weeks (Form 1 aggregate, Form 2 aggregate). Because only a single version of the *Content Analysis-Constructed Response* probe was administered each testing session, our aggregate of the two scores produced a single *Content Analysis-Constructed Response* score, representing the average of the student's scores on the *Content Analysis-Constructed Response* probe administered in each of the data collection sessions.

Table 12. Correlations between grade-based measures, teacher ratings, and aggregated probe scores

		Overall GPA	Grade in Algebra		Teacher Rating	
Basic Skills	2 min.	NS	.27	(1 NS, .24 -.30)	.33	(.30 - .38)
	3 min.	NS	.24	(.21 - .27)	.34	(.30 - .36)
	4 min.	NS	.23	(.21 - .25)	.33	(.29 - .35)
Algebra Foundations	4 min.	NS	.35	(2 NS, .27 -.43)	NS	
	5 min.	NS	.34	(2 NS, .25 - .42)	NS	
	6 min.	NS	.35	(2 NS, .28 - .41)	NS	(3 NS; .21)
Content Analysis-Constructed Response		NS	.23		.39	
Translations	Correct	NS	.25	(2 NS, .23 - .26)	NS	
	C – I	NS	.41	(.37 - .45)	.29	(.28 - .30)
	1/3	NS	.32	(.27 - .40)	.22	(2 NS; .21 - .22)
Content Analysis-Multiple Choice	Correct	NS	NS	(3 NS; .31)	.27	(.22 - .32)
	C – I	NS	.29	(1 NS, .23 - .39)	.29	(.22 - .36)
	1/3	NS	NS	(3 NS; .34)	.28	(.22 - .34)

Aggregating scores across two probes had minimal impact on the criterion validity of the *Basic Skills, Algebra Foundations*, or the *Content Analysis-Multiple Choice* measures with regard to grade-based measures and teacher ratings (Table 12). Slight improvements in criterion validity for the two corrected scores were obtained for the *Translations* measure. A mixed pattern of results was obtained for the *Content Analysis-Constructed Response* measure. The correlation with algebra grades decreased (from .40 to .23), while the correlation with teacher ratings increased from .37 to .39. With regard to relations between the aggregated scores on the algebra measures and the Iowa Test of Educational Development (ITED; Table 13), a similar pattern of limited improvement was obtained. For the *Content Analysis-Constructed Response* measure, the correlations using the aggregated scores were lower than the average of correlations based on single scores.

A different pattern of results emerged when students' aggregated scores on the algebra probes were correlated with their scores on the Iowa Algebra Aptitude Test (IAAT; Table 14). For this criterion measure, aggregating scores had the effect of increasing the strength of the

Table 13. Correlations between ITED subtests and aggregated probe scores

		ITED Scores		
		Con/Prob	Comp	Reading Total
Basic Skills	2 min.	.25 (2 NS, .22 - .27)	.29 (.23 - .32)	NS
	3 min.	.23 (2 NS, .22 -.24)	.27 (.22 - .29)	NS
	4 min.	.24 (2 NS, .23 - .25)	.28 (.23 - .32)	NS
Algebra Foundations	4 min.	.37 (2 NS, .35 - .39)	.23 (1 NS, .22 - .25)	.33 (1 NS, .27 - .40)
	5 min.	.36 (1 NS, .26 - .44)	.26 (.24 - .28)	.34 (.27 - .43)
	6 min.	.31 (.23 - .39)	.28 (1 NS, .26 - .29)	.33 (.30 - .38)
Content Analysis-Constructed Response		.27	.27	.30
Translations	Correct	NS (3 NS; .24)	NS	NS
	C – I	.24 (1 NS, .22 - .27)	.24 (2 NS, .22 - .25)	NS
	1/3	NS (3 NS; .24)	NS	NS
Content Analysis-Multiple Choice	Correct	.32 (2 NS, .28 - .35)	.24 (2 NS, .21 - .27)	NS (3 NS; .20)
	C – I	.30 (2 NS, .28 - .31)	NS (3 NS; .23)	NS
	1/3	.32 (2 NS, .29 - .34)	NS (3 NS; .26)	NS

relation with the IAAT. The *Content Analysis-Constructed Response* and the *Algebra Foundations* measure showed moderate relations (.47 to .55 range). The correlation for the *Content Analysis-Constructed Response* measure was in a similar range (.52), but this coefficient was lower than the average for the single probes (.56). The *Basic Skills* and *Translations* measures both had correlation coefficients in the .3 range.

Table 14. Correlations between the IAAT subtest and total test scores and aggregated probe scores

	Total	Part A	Part B	Part C	Part D
Basic Skills					
2 min.	.42 (.37 - .44)	NS	.35 (.32 - .39)	.32 (.27 - .37)	.33 (1 NS, .28 - .37)
3 min.	.39 (.33 - .42)	NS	.33 (.27 - .38)	.29 (.26 - .33)	.32 (1 NS, .24 - .38)
4 min.	.40 (.32 - .37)	NS	.33 (.27 - .37)	.31 (.29 - .33)	.34 (1 NS, .24 - .41)
Algebra Foundations					
4 min.	.48 (.36 - .57)	.30 (2 NS, .27 - .32)	.39 (.29 - .48)	.40 (.36 - .45)	.35 (2 NS, .33 - .37)
5 min.	.54 (.49 - .57)	.30 (.25 - .33)	.44 (.41 - .46)	.43 (.41 - .48)	.34 (1 NS, .26 - .38)
6 min.	.55 (.53 - .57)	.32 (1 NS; .29 - .35)	.44 (.40 - .48)	.46 (.43 - .48)	.30 (.28 - .35)
Content analysis-Constructed Response					
	.52	.29	.46	.41	.36
Translations					
Correct	.26 (2 NS, .24 - .27)	.22 (2 NS, .22 - .22)	NS	.24 (1 NS, .22 - .28)	NS
C – I	.36 (.27 - .40)	.32 (1 NS, .24 - .36)	.29 (1 NS, .28 - .30)	.26 (1 NS, .22 - .30)	.30 (2 NS, .29 - .30)
1/3	.32 (.26 - .34)	.29 (1 NS, .24 - .32)	.26 (1 NS; .24 - .28)	.26 (.23 - .29)	NS
Content Analysis-Multiple Choice					
Correct	.47 (.38 - .54)	.31 (.29 - .34)	.35 (.26 - .42)	.31 (.23 - .36)	.35 (.30 - .38)
C – I	.48 (.47 - .53)	.32 (.30 - .33)	.37 (.30 - .42)	.34 (.29 - .38)	.33 (.27 - .38)
1/3	.49 (.41 - .55)	.32 (.29 - .35)	.37 (.28 - .43)	.33 (.26 - .37)	.34 (.29 - .39)

Summary and Considerations for Future Research

The purpose of this study was to examine the technical adequacy of five potential measures of algebra proficiency. One hundred three students in grades nine to twelve from two different Iowa school districts participated in the study; 14 of these students were receiving special education services. The data were gathered from October through December 2004, when students were in the second half of a semester-long algebra course (both districts use block scheduling). Over four weeks of probe data collection, students completed two forms of a *Basic Skills* probe, two forms of an *Algebra Concepts* probe, a single form of a *Content Analysis-Constructed Response* probe, two forms of a *Translations* probe, and two forms of a *Content Analysis-Multiple Choice* probe. The testing sessions were spaced one week apart and were preceded two weeks earlier by the administration of the Iowa Algebra Aptitude Test (IAAT). Data collected on additional criterion variables included students' grades in school and in algebra, teachers' ratings of students' proficiency in algebra, scores on the district's standardized achievement test (Iowa Tests of Educational Development) and scores on the IAAT. This summary reviews the major findings with respect to score distributions, reliability, criterion validity, barriers encountered, and issues for future research.

Distributions

Of the five measures explored in this study, only one demonstrated evidence of problems with the distribution of student scores. On the *Translations* probe, the mean number of incorrect responses was nearly as large as the number correct responses, suggesting a high rate of guessing. The remaining probes did not show evidence of floor or ceiling effects. In addition, none of the other four probes produced an over-abundance of 0 scores; all demonstrated sufficient room for students to improve their performance and continue to demonstrate improvement over the remainder of the course.

Reliability

The reliability of individual probe scores was slightly higher than the results reported in Technical Report 6 (Foegen & Olson, 2005), but continued to fall short of standard benchmarks. Coefficients for single probe reliability were in the .4 to .9 range. Inconsistent results were obtained when we examined alternate scoring methods for correcting the raw scores for guessing, with no clear improvements favoring either of the correction methods for single probes. When scores from two probes were aggregated to increase the stability of the estimate of student performance, the estimates for test-retest reliability increased for four probes (students completed only a single form of the *Content Analysis-Constructed Response* measure in each session, so it was not possible to examine the effects of aggregation on this measure). With regard to alternate form reliability, the use of aggregated scores produced reliability coefficients that were similar or slightly improved for the *Basic Skills* and *Algebra Foundations* measures. For the *Translations* measure and the *Content Analysis-Multiple Choice* probes, the results were mixed, with some coefficients increasing, while others remained similar or decreased. Neither the single probes, nor the aggregated scores met the expected .80 levels necessary to use the measures for educational decision-making.

Validity

To examine the criterion validity of the measures, we computed correlations between single and aggregated probe scores with grade-based measures (grade in algebra and overall GPA), teacher ratings, standardized test scores, and scores on the Iowa Algebra Aptitude Test (IAAT). Criterion validity coefficients for single probes were generally in the low range, from .2 to .4. The strongest coefficients were obtained for the *Content Analysis-Constructed Response* measure, regardless of the criterion variable used. The *Content Analysis-Multiple Choice* and *Algebra Foundations* measures had the next strongest relations with the criterion measures, followed by the *Basic Skills* and *Translations* measures. Aggregating scores produced only slight improvements in the criterion validity coefficients for both measures.

Considerations for Future Research

Several issues arose during this study that should be addressed in future research. First, we observed that students' scores on the first administration of a task were often substantially lower than their scores on subsequent tasks. In the future, it may be useful to incorporate a practice task that allows students to become familiar with the format of the problems and thereby reduces the 'learning curve' effect we observed between the first and subsequent administrations.

Second, we identified several concerns related to the scoring of the *Content Analysis-Constructed Response* measure. Although this measure had the highest levels of criterion validity across the five measures, it had the lowest levels of interscorer agreement. In addition, we found that the measure was extremely time consuming to score, which limits the practicality of the measure for classroom teachers. We were surprised by the differences in the results obtained for this measure relative to its multiple choice 'sibling,' the *Content Analysis-Multiple Choice* measure. The latter measure was dramatically easier to score efficiently and consistently, so we anticipate exploring further refinements to this measure in an effort to increase the technical adequacy of this measure.

Third, we identified one potential problem in the design of the *Content Analysis-Multiple Choice* measure. The probe was developed by generating problems associated with one to three key concepts or skills from each chapter of the textbook. On the probe, each chapter was represented by one or two questions. In situations where there were more key concepts than questions, the specific skill or concept sampled varied from one form of a probe to another. This design characteristic may have introduced additional variance to students' scores.

Finally, we occasionally encountered instances in which students did not appear to be giving their best effort in responding to the probes. This may have been due in part to the informed consent process, during which students were informed that their scores on the probes would not affect their grades in algebra class. It may be valuable to explore options with teachers that would create an incentive for students to do their best work on the probes.

Future research involving the algebra progress monitoring measures should examine the following issues:

- Incorporating a 'practice probe' each time a new probe format is introduced to allow students to become familiar with the format and content of the measure
- Revising the design template for the *Content Analysis-Multiple Choice* measure so that alternate forms are assessing parallel content
- Refining scoring rubrics for the *Content Analysis-Multiple Choice* measure to further increase interscorer agreement

- Exploring (with participating teachers) means of encouraging students to do their best work on the probes

References

Foegen, A. (2000). Technical adequacy of general outcome measures for middle school mathematics. *Diagnostique, 25*, 175-203.

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

Foegen, A., & Lind, L. S. (2004). *Reliability and criterion validity for three potential algebra measures* (Tech. Rep. No. 2). Ames, IA: Iowa State University, Project AAIMS.

Foegen, A., & Olson, J. (2005). *Reliability and criterion validity for two algebra measures: Translations and Content Analysis-Multiple Choice* (Tech. Rep. No. 6). Ames, IA: Iowa State University, Project AAIMS.

Fuchs, L. S., Deno, S. L., & Marston, D. (1983). Improving the reliability of curriculum-based measures of academic skills for psychoeducational decision making. *Diagnostique, 8*, 135-149.

Lappan, G., Fey, J. T., Fitzgerald, W. M, Friel, S. N., & Phillips, E. D. (2004). *Prentice Hall Connected Mathematics*. Upper Saddle River, NJ: Pearson Prentice Hall.

Moses, R. P., & Cobb, C. E. (2002). *Radical equations: Civil rights from Mississippi to the Algebra Project*. Boston, MA: Beacon.

APPENDIX

Basic Skills – Form 1

Basic Skills – Form 2

Algebra Foundations – Form 1

Algebra Foundations – Form 1

Content Analysis-Constructed Response – Form 1

Content Analysis-Constructed Response – Form 1

Translations – Form 1

Translations – Form 2

Content Analysis-Multiple Choice– Form 1

Content Analysis-Multiple Choice – Form 2

Standard Directions for Administration

Teacher Rating Form

$9 + a = 15$ $a =$	Simplify: $y^2 + y - 4y + 3y^2$
Simplify: $2x + 4 + 3x + 5$	$12 - 6 = e$ $e =$
$12 - e = 4$ $e =$	Simplify: $12 + (-8) + 3$
Simplify: $(-5) + (-4) - 1$	$\frac{z}{5} = 5$ $z =$
$6 \bullet 9 = d$ $d =$	Simplify: $4d + 7d - 9 + 2$
Simplify: $4(3 + s) - 7$	$\frac{54}{w} = 9$ $w =$
$2s = 16$ $s =$	Simplify: $8 - (-6) - 4$
Simplify: $4 - (-2) + 8$	$6 + 7 = v$ $v =$
$8m = 72$ $m =$	Simplify: $6r - 5 - 2r + 6$
Simplify: $3(c + 2) - 2c$	$f - 7 = 3$ $f =$
$\frac{63}{c} = 9$ $c =$	Simplify: $-5 + 6 - 6$
Simplify: $17 + 6d + 2d - 9$	$r - 4 = 4$ $r =$
$y + 4 = 11$ $y =$	Simplify: $w - w(4+5) - 6$
Simplify: $8m - 3(m - 2)$	$\frac{36}{6} = s$ $s =$
$6 \bullet 3 = k$ $k =$	Simplify: $3z - 8z + 2 + 9$

$3 \cdot 8 = m$ $m =$
Simplify: $9 + (-3) - 8$
$b + 8 = 8$ $b =$
Simplify: $4 - 7b + 5(b - 1)$
$\frac{18}{g} = 6$ $g =$
Simplify: $-2 + (-5) + (-8)$
$11 - n = 6$ $n =$
Simplify: $-3w^2 + 5w^2 - 5 + 12$
$f + 7 = 15$ $f =$
Simplify: $s + 2(s - 5) - 3$
$4r = 28$ $r =$
Simplify: $5 - 2b + 4(b + 3)$
$2 + t = 7$ $t =$
Simplify: $1 - 9 + (-2)$
$5q = 30$ $q =$

Simplify: $3(u + 3) - 2u + 5$
$\frac{28}{4} = d$ $d =$
Simplify: $7b - 4 - 3 - 2b$
$18 - 9 = k$ $k =$
Simplify: $-9 + 3 + 8$
$\frac{h}{6} = 8$ $h =$
Simplify: $11 - 8g - 2 - 4g$
$4 + 7 = x$ $x =$
Simplify: $14 - 7 + (-3)$
$5 + z = 13$ $z =$
Simplify: $16 + 2(t - 4) - 3t$
$5 \cdot 7 = j$ $j =$
Simplify: $9 - 4(v - 2)$
$12 - d = 4$ $d =$
Simplify: $b + b + 2b$

$16 - p = 7$ $p =$	Simplify: $5q - 7 - 2 - 3q$
Simplify: $6 - 2(v - 7)$	$3t = 21$ $t =$
$9 \bullet 5 = a$ $a =$	Simplify: $8 - 3g + 6(g + 2)$
Simplify: $z + z + 3z$	$2 + a = 8$ $a =$
$8 + 3 = t$ $t =$	Simplify: $5 - 3 + (-8)$
Simplify: $16 - 5 + (-3)$	$13 - n = 5$ $n =$
$9 + b = 14$ $b =$	Simplify: $-6m^2 + 2m^2 - 8 + 9$
Simplify: $15 + 3(y - 6) - 3y$	$h + 8 = 11$ $h =$
$12 - 5 = j$ $j =$	Simplify: $p + 3(p - 6) - 4$
Simplify: $-7 + 11 + 2$	$a + 5 = 5$ $a =$
$\frac{h}{8} = 7$ $h =$	Simplify: $6 - 9c + 7(c - 1)$
Simplify: $13 - 5v - 4 - 2v$	$\frac{21}{v} = 3$ $v =$
$8e = 40$ $e =$	Simplify: $-1 + (-4) + (-7)$
Simplify: $5(u + 8) - 3u + 9$	$\frac{49}{7} = w$ $w =$
$\frac{63}{9} = s$ $s =$	Simplify: $4z - 8z + 4 + 5$

$6 \bullet 8 = y$
$y =$

Simplify:
$9 + (-4) - 8$

$e - 5 = 4$
$e =$

Simplify:
$-7 + 9 - 9$

$r - 2 = 6$
$r =$

Simplify:
$f - f(2 + 8) - 7$

$\frac{42}{s} = 6$
$s =$

Simplify:
$7 - (-4) - 2$

$7 + 5 = j$
$j =$

Simplify:
$9k - 4 - 2k + 3$

$18 - 9 = p$
$p =$

Simplify:
$13 + (-7) + 5$

$\frac{d}{9} = 9$
$d =$

Simplify:
$4d + 7d - 9 + 2$

$x + 6 = 14$
$x =$

Simplify:
$6c - 4(c - 7)$

$7 \bullet 3 = k$
$k =$

Simplify:
$y^2 + y - 3y + 9y^2$

$9q = 81$
$q =$

Simplify:
$4(c + 8) - 3c$

$\frac{56}{a} = 7$
$a =$

Simplify:
$10 + 6d + 4d - 8$

$8 \bullet 6 = w$
$w =$

Simplify:
$5(9 + s) - 4$

$2s = 18$
$s =$

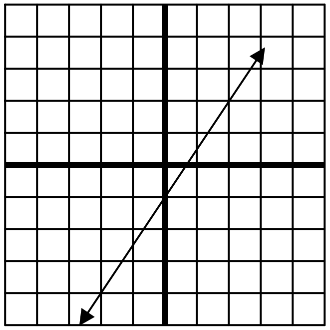
Simplify:
$7 - (-3) + 9$

$7 + b = 12$
$b =$

Simplify:
$3m + 5 + 4m + 2$

$14 - e = 9$
$e =$

Simplify:
$-6 + (-3) - 2$



What is the slope?

What is the *y* intercept?

Fill in the empty box:

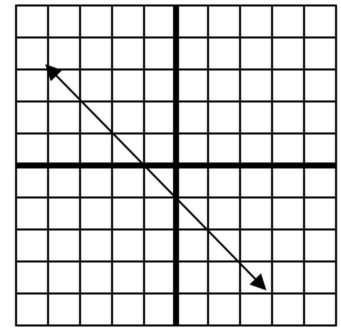
s	$3s$
6	18
7	21
8	
9	27

Fill in the empty box:

n	$4n+7$
1	11
2	15
3	
4	23

Fill in the empty box:

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



What is the slope?

What is the *y* intercept?

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

6 • 4 + 1 = _________

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

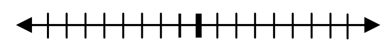
Solve:
$n + 3 = 8$
n = _________

Evaluate $4b + 2$ when
$b = 1$ _________ and when
$b = 3$ _________

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

(-2) • (-4) = _________

Graph the expression
$m > 6$



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

8 ÷ 2 + 4 • 3 = _________

2^3 = _________

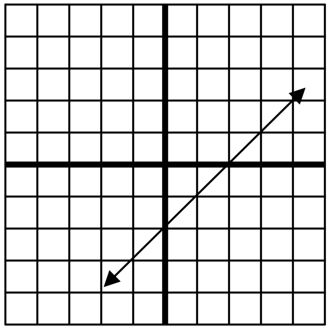
Write the expression for this phrase:
9 times a number

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

Evaluate $2x + y$ when
$x = 2$ and $y = 3$
____

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

Simplify:
$6 – 2(b – 4)$



What is the slope?

What is the *y* intercept?

Fill in the empty box:

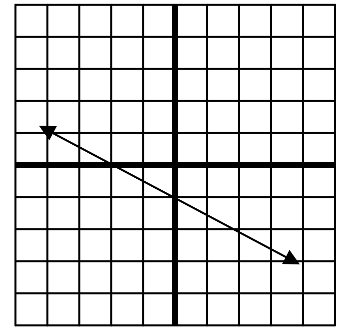
n	
6	4
9	6
12	8
15	10

Fill in the empty box:

t	*t* - 7
5	-2
6	-1
8	1
10	

Fill in the empty box:

w	
4	11
6	17
8	23
10	29



What is the slope?

What is the *y* intercept?

Write a word phrase for this expression:
$x \div 4$

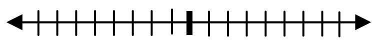
$(-16) \div (-4) =$ _________

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

Solve:
$3x = 27$
$x =$ _________

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

Graph the expression $p \leq -3$



Simplify:
$9x - 3 - 4x + 9 =$

Solve:
$24 \div x = 6$
$x =$ _________

Evaluate $8g - 4$ when
$g = 2$ __________
$g = 4$ _________

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

$9 \cdot 4 - 6 =$ _________

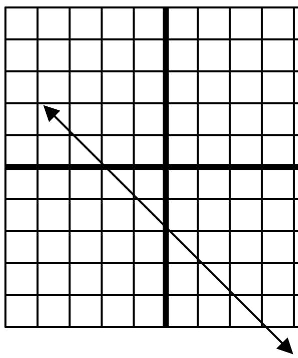
Simplify:
$12n - 5 - 7n + 3$

Write a word phrase for this expression:
4 times a number

$(-3)(9 - 7) =$ __________

$\sqrt{36} =$ _________

Simplify:
$3(6 - 1) + 2(-4 + 4)$



What is the slope?

What is the y intercept?

Fill in the empty box:

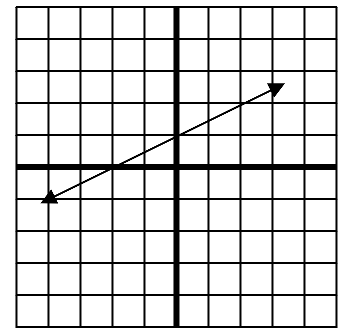
t	$4t + 1$
2	9
4	17
5	
10	41

Fill in the empty box:

y	
5	1
10	2
15	3
50	10

Fill in the empty box:

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



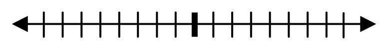
What is the slope?

What is the y intercept?

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

Solve:
$49 \div n = 7$
$n =$

Graph the expression $t < 2$



$18 \div 3 + 6 \bullet 4 =$

Evaluate $6s + 4$ when
$s = 4$ _________ and when
$s = 6$ _________

Write a word phrase for this expression:
$c + 18$

$(-6) \bullet (-8) =$ _________

If $y > 3$, two possible values for y are _________ and _________

$(-5)(8 - 6) =$ _________

$\sqrt{49} =$ _________

Simplify:
$8t + (3t - t)$

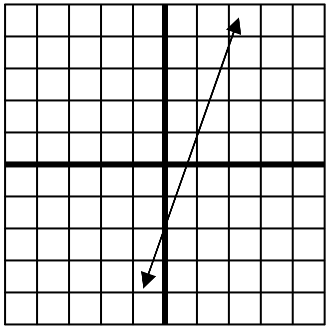
Solve:
$9x = 45$
$x =$ _________

$4 \bullet 5 - 2 + 6 =$ _________

If $2a + 4 \geq 12$, two possible values for a are _________ and _________

Write a word phrase for this expression:
$j - 12$

Simplify:
$2 + 2 \bullet 4 - 4$



What is the slope?

What is the y intercept?

Fill in the empty box:

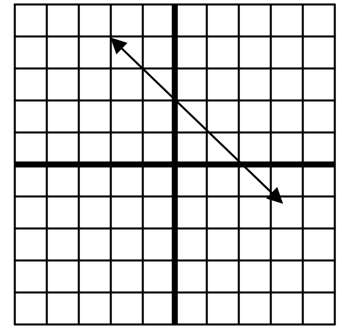
x	
10	5
20	10
50	25
100	50

Fill in the empty box:

n	$2n$-3
3	3
5	7
7	11
9	

Fill in the empty box:

y	
1	8
2	9
5	12
10	17



What is the slope?

What is the y intercept?

Simplify:
$12n - 7 + 3n + 4$

Solve:
$b + 7 = 16$
$b =$

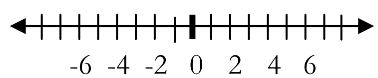
Write the expression for this phrase: *10 less than 3 times a number*

Write a word phrase for this expression:
$5t$

$-4(8 + 2) =$ _________

$3^3 =$ _________

Graph the expression $p \geq -5$



Evaluate $4a - b$ when
$a = 3$ and $b = 4$
____

Solve $18 - n = 12$
$n =$ _________

Simplify:
$9g + (2g - g)$

Write the expression for this phrase: *6 more than 4 times a number*

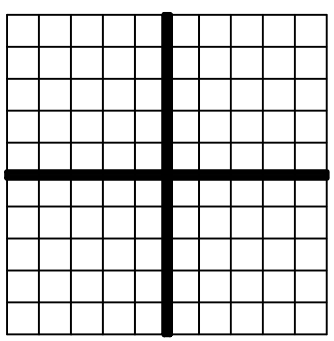
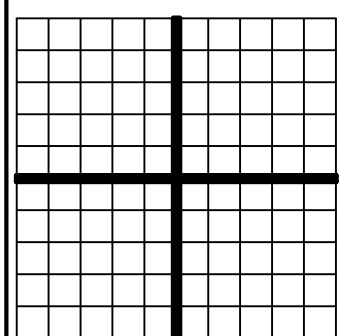
$8 \div 2 + 6 \cdot 2 =$

Write a word phrase for this expression:
$\frac{18}{b}$

Evaluate $5x - 4$ when
$x = 4$ _________
$x = 8$ _________

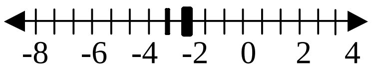
Write the expression for this phrase:
3 times a number

Simplify:
$5(m + 2) - 3m$

Evaluate $5y + x^2$ when $y = 3$ and $x = 2$	Simplify: $8(-4)(b)(b)(b)$	Solve: $\frac{x}{2} = (-6)$	Solve: $2t - 4 = 6$
Graph $x = 3$ 	Graph the equation: $y + 2x = 4$ 	Write the equation in slope-intercept form: $m = (-8)$ $b = 4$	Write the equation of a line that passes through (3, -1) and has a slope = -2. Use slope-intercept form.

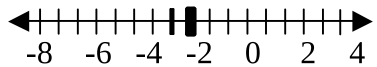
Solve and graph:

$x < (-2)$ or $3x - 5 > 1$



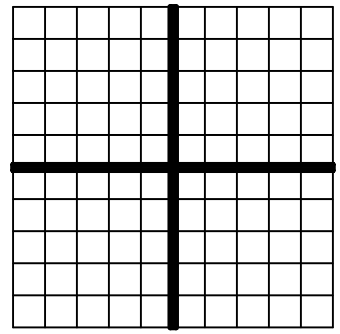
Solve and graph:

$(-5) + x \geq 1$



Graph the linear system:

$y = 2x - 3$ $-y = 2x - 1$



Simplify the expression:

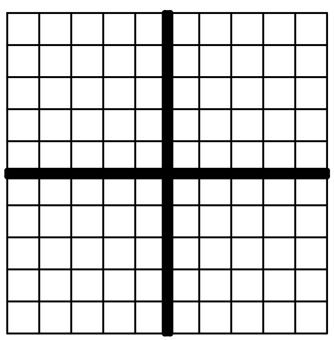
$$\frac{x^3}{xy^4} \bullet \frac{y^5}{x^5}$$

Circle the function that matches the graph:

$y = \left(-x^2\right) - 2x + 3$

$y = \left(-3x^2\right) - x + 2$

$y = 2x^2 + x - 3$



Subtract:

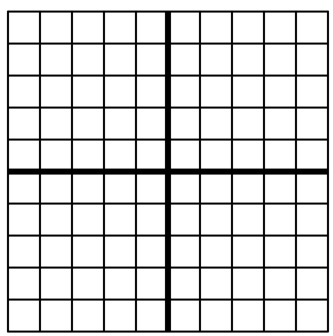
$(5t^2 - 9t + 1)$ - $(8t + 13)$

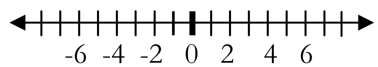
Simplify the expression:

$$\frac{x^2 - x - 6}{x^2 - 4}$$

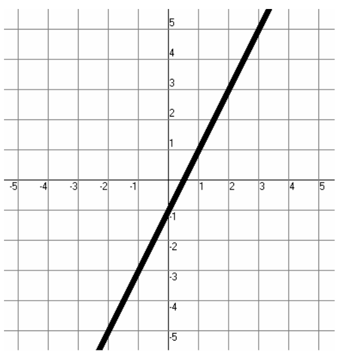
Solve:

$$\sqrt{2m+3} - 6 = 4$$

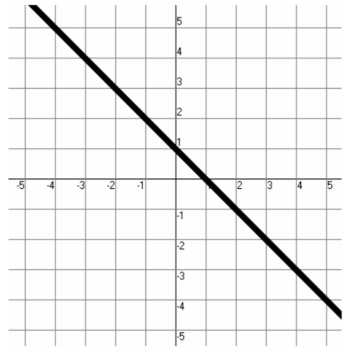
Add parentheses to make the expression true: $4 + 6 \div 2 = 5$	Rewrite without parentheses: $2(a + 3) - 2(a - 1)$	Solve: $6x - 9 = 10x + 3$	Solve: $5m = 25$
Find the slope of a line through (2, 4), (5, 0)	Graph the equation: $y = 2x$ 	Write the equation of a line through (-2, 5), (2, 4)	Write the equation in slope-intercept form: $m = 2$, $b = 6$

Solve: $\| 2x \| > 14$	Solve and graph: $x < 2$ and $x \geq (-3)$  	Solve the linear system: $(-6x) + 3y = (-6)$ $2x + 6y = 30$	Simplify the expression: $(mn)^2 \bullet n^4$
Evaluate the expression: $\sqrt{100}$	Solve: $(6x - 5)(x + 2) = 0$	Write the product in simplest form: $\frac{6x^2}{8x} \bullet \frac{-4x^3}{2x^2}$	Solve by completing the square: $x^2 - 2x = 2$

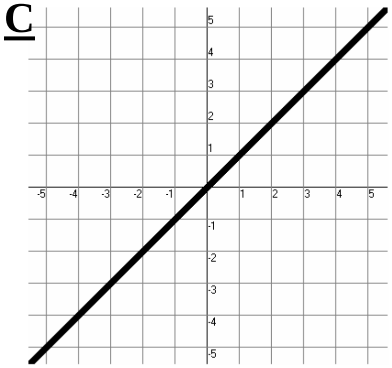
A



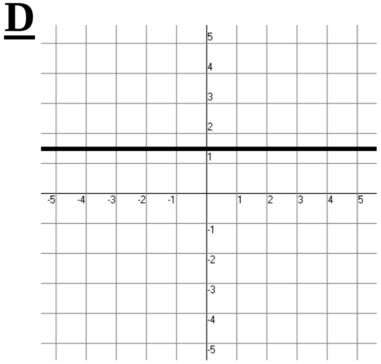
B



C



D



$y = x$ ____	$y = 2x - 1$ ____	$2y = 4x - 2$ ____	$y = 1.5$ ____	$y = -x + 1$ ____

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

____

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

____

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

____

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

____

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

____

Mark needs to find half the width of pieces of pipe he is cutting to make a soccer goal. The width of the pipe is 3 inches. He made this graph to show the relationship between the length and the width of the pieces he will cut.	____
Every day that Cindy waters the garden, she earns a dollar. She made this graph to show the relationship between the number of days she waters the garden and the number of dollars she will earn.	____
Joe has one dollar in his wallet. He made this graph to show the relationship between the number of dollars he borrows from his friends for lunch and the total amount of money he has or owes.	____
The class earns $2 for each magazine subscription sold in the fund-raiser. A $1 fee per student is charged for a processing fee. Cindy made this graph to show the relationship between the number of magazines sold and the profit.	____
The flood waters are receding at a rate of 1 foot per day. The river is currently at 1 foot above flood stage. Tom made this graph to show the relationship between the number of days and the height of the river compared to flood stage.	____

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

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

____

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

____

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

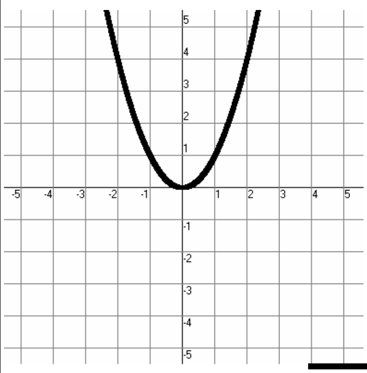
____

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

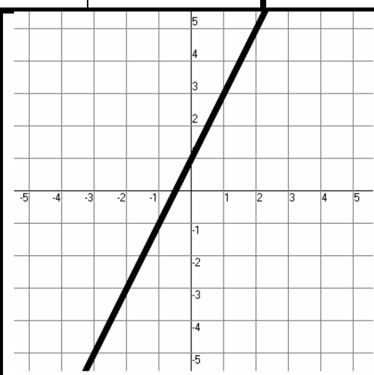
____

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

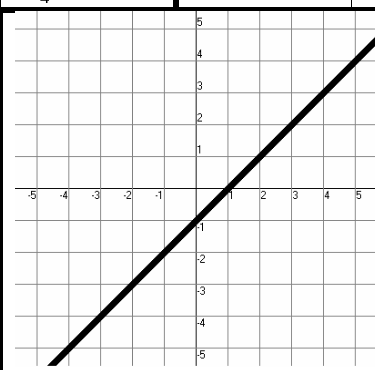
____



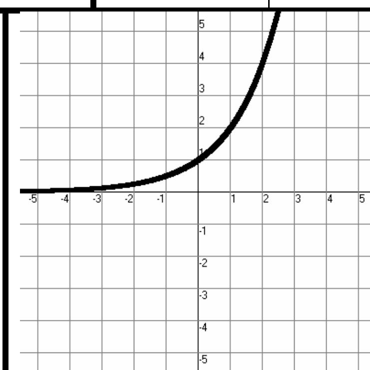
____



____



____



____

Mr. Jones is going to give a true/false test. He wrote this equation to show the number of possible answer combinations his students can give on the test. ____

Sue wrote this equation to figure out how many inches of wire she needs for a bracelet. Each bracelet uses two strands and she needs to add an extra inch to make a hood to fasten the bracelet. ____

Sam's allowance changes every year. Each month his mom pays him a dollar for each year he has lived, multiplied by his age. Sam wrote this equation to figure out his allowance. ____

Every time Hans delivers newspapers, he keeps one for his family. Hans wrote this equation to show how many newspapers he delivers to families on his route. ____

Tim's washing machine 'eats' socks. The first time he lost one sock in the wash. Now, every time he washes a load of clothes, he loses two socks. Tim wrote this equation to figure out how many socks he is losing. ____

A

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

B

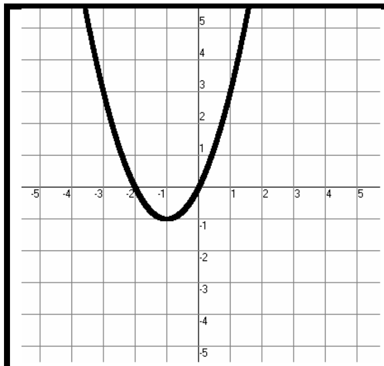
x	y
4	2
2	1
0	0
-2	-1
-4	-2

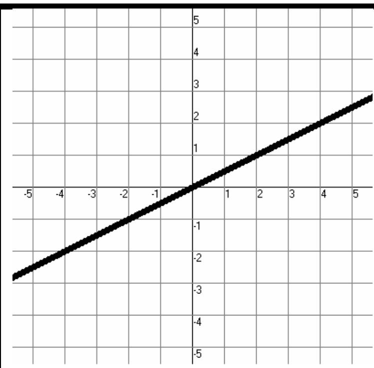
C

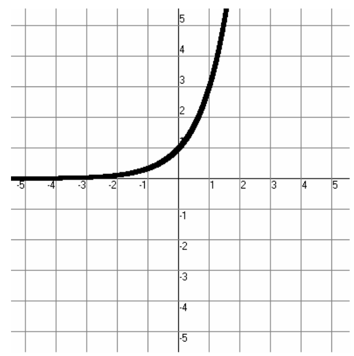
x	y
2	9
1	3
0	1
-1	$\frac{1}{3}$
-2	$\frac{1}{9}$

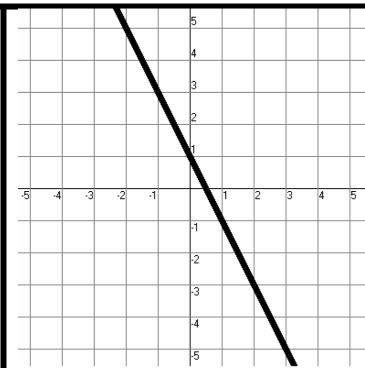
D

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


______


______

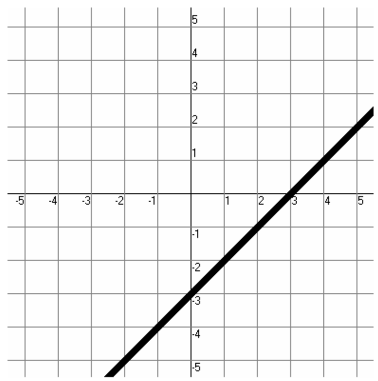

______


______

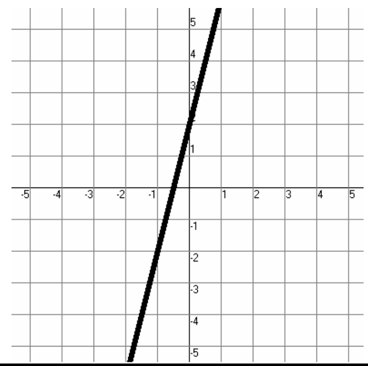
$y = -2x + 1$ ______	$y = x^2 + 2x$ ______	$y = 3^x$ ______	$y = x(x + 2)$ ______	$y = \frac{1}{2}x$ ______

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

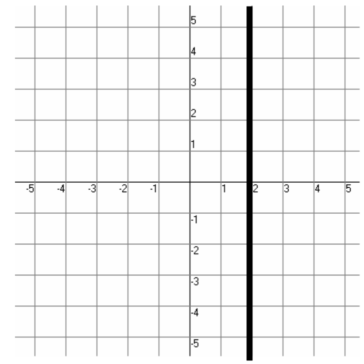
A



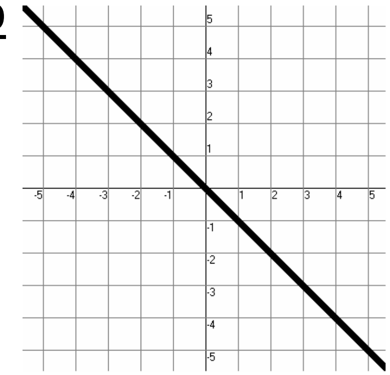
B



C



D



$3y = 3x - 9$ ____	$y = 4x + 2$ ____	$y = x - 3$ ____	$y = -x$ ____	$x = 2$ ____

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

____

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

____

x	*y*
2	-2
1	-1
0	0
-1	1
-2	2

____

x	*y*
4	18
2	10
0	2
-2	-6
-4	-14

____

x	*y*
4	1
2	-1
0	-3
-2	-5
-4	-7

____

Tim is collecting state quarters for his state. He started his collection with two quarters. He wants to trade in some dollar bills for quarters. Tim made this graph to show how many quarters he'll have after the trade.	____
Leah is three years younger than her sister. She made this graph to show the relationship between their ages.	____
Every time he gets home after curfew, he loses a chance to use the car. Joel made this graph to show the relationship between breaking curfew and his chances to use the car.	____
Sam is planning a basketball tournament. He made this graph to show the relationship between the number of teams in the championship game and the total number of teams in the tournament.	____
Teresa has taken four quizzes and gotten the same score on each one. She also has two extra credit points. Teresa made this graph to show how her total quiz points would be related to the score she gets on each quiz.	____

A	**B**	**C**	**D**
$y = 16(.5)^x$	$y = -2x - 1$	$y = x + 1$	$y = x^2 - 1$

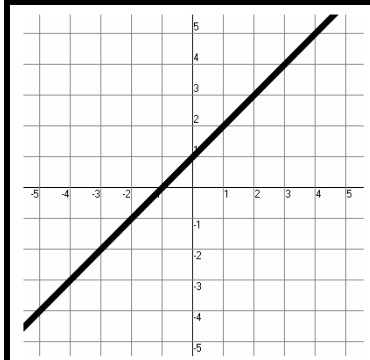
x	y
2	-5
1	-3
0	-1
-1	1
-2	3

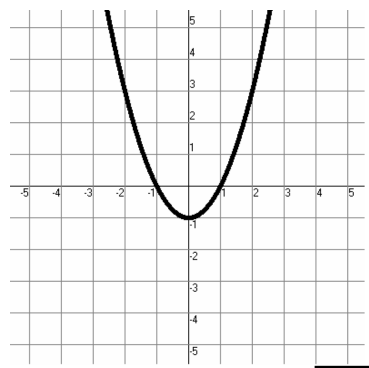
x	y
4	15
2	3
0	-1
-2	3
-4	15

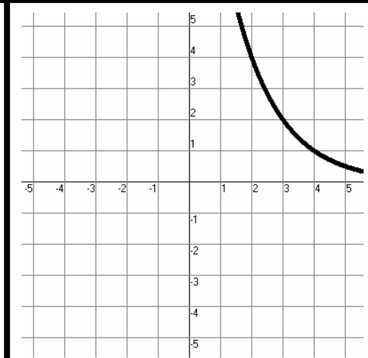
x	y
2	4
1	8
0	16
-1	32
-2	64

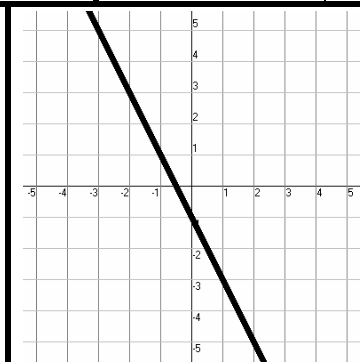
x	y
2	3
1	2
0	1
-1	0
-2	-1

x	y
4	-9
2	-5
0	-1
-2	3
-4	7









Pat is organizing the brackets for the doubles tennis tournament. Sixteen teams have entered. Pat wrote this equation to show how many teams will be left after each of the rounds.

LeRoy needs to buy tile for a square room. The tiles come in 1-foot squares. There is a post in the middle of the room that is the same size as one tile. LeRoy wrote this equation to find how many tiles he will need.

Elaine's mom gives her a list of chores to do each week. Before the week is over, she always finds one more thing that Elaine needs to do. Elaine wrote this equation to show the number of chores she does each week.

When Maria eats hot lunch, it costs two dollars. She already owes her sister a dollar. Maria wrote this equation to find out how much less money she'll have each time she eats hot lunch.

Ryan has a stool that is one foot tall. He wrote this equation to find the height of any person who stands on the scale.

A

x	y
4	14
3	6
2	2
1	0
0	-1

B

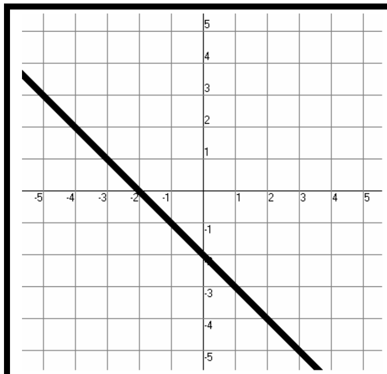
x	y
4	8
2	4
0	0
-2	-4
-4	-8

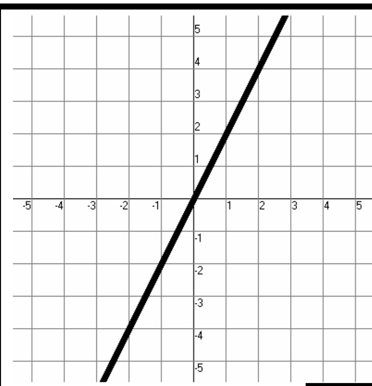
C

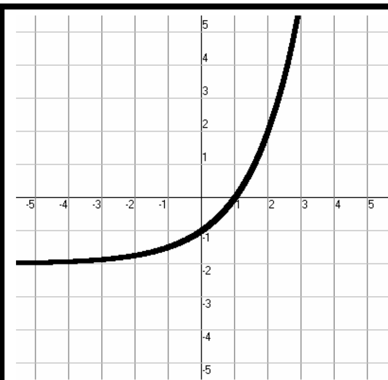
x	y
2	10
1	5
0	2
-1	1
-2	2

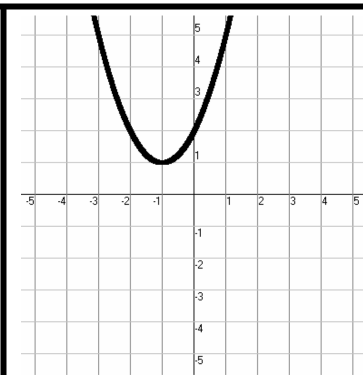
D

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









$y = x^2 + 2x + 2$ ____	$y = - x - 2$ ____	$y = 2x$ ____	$y = 2^x - 2$ ____	$y + 2 = 2^x$ ____

Bryan's dad will match his donation to the animal shelter. Brian made this table to show the relationship between how much he gives and his total donation to the shelter. ____

At the teachers' cookie swap, each teacher brings one cookie for all the teachers. The principal brings two cookies for each teacher. The cooks donate two cookies left from lunch. This table shows the number of teachers and cookies. ____

The class is planting trees for Earth Day. Each hole needs to be dug two feet deeper than the height of the root ball. This table shows the relationship between the root ball's height and the level of the ground. ____

Chris learned that a pair of mice will produce one litter of two baby mice and that when each baby matures, it will do the same. Chris made this table to show the relationship between the generations and the total mice if the original two mice die. ____

Jean changed jobs and doubled her hourly pay rate. This table shows the relationship between Jean's old and new hourly pay rates. ____

Evaluate $b^2 - a^2$ when $a = 4$ and $b = 5$

a) 21
b) 1
c) 11
d) 9

Rewrite this expression without parentheses:
(-5) (4 – y)

a) 9 - y
b) -20 + 5y
c) -1 – 5y
d) -20 – 5y

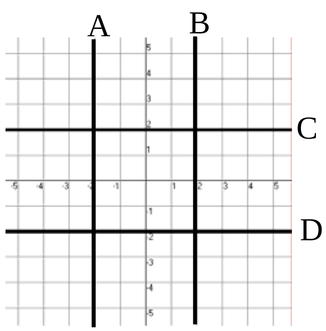
Solve:
$2t - 5 = 7$

a) $\frac{1}{2}$
b) 6
c) 1
d) 2

Solve:
$\frac{y}{3} = 4$

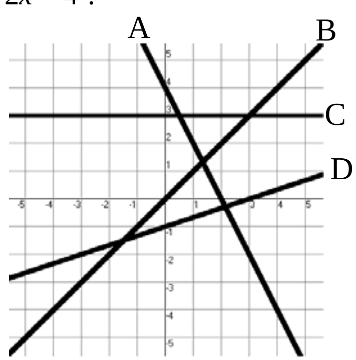
a) -10
b) 7
c) $\frac{4}{3}$
d) 12

Which line on the graph is
y = 2?



a) Line A
b) Line B
c) Line C
d) Line D

Which line on the graph is
$y + 2x = 4$?



a) Line A
b) Line B
c) Line C
d) Line D

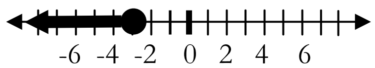
Write the equation in slope-intercept form:
$m = (\frac{1}{2})$ $b = 3$

a) $y = 2x + 3$
b) $y = 3x + \frac{1}{2}$
c) $x = \frac{1}{2}y - 3$
d) $y = \frac{1}{2}x + 3$

Rewrite this equation in standard form using integer coefficients.
$-4y + \frac{1}{2}x = 2$

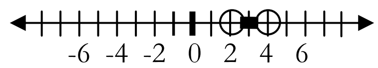
a) $-8y + 2x = 4$
b) $x - 8y = 4$
c) $y = 4x + 8$
d) $4y - 2x = 4$

This graph shows the solution for which equation?



a) $x > -3$
b) $2x \le -6$
c) $-3x > 9$
d) $3x \ge 9$

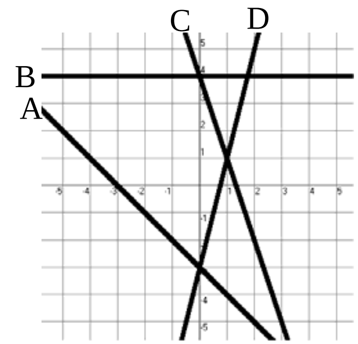
This graph shows the solution for which equation?



a) $3x > 6$ or $2x < 2$
b) $2 < 4x - 6 < 10$
c) $2 < |x| < 4$
d) $2x < 6$

Circle the TWO lines that show the solution to this linear system:

$4x - y = 3$
$3x + y = 4$



a) Line A
b) Line B
c) Line C
d) Line D

Evaluate the expression:

4^{-2}

a) -16
c) $\frac{1}{16}$
b) $\frac{1}{8}$
d) -8

Simplify

$\sqrt{32}$

a) $4\sqrt{2}$
b) $8\sqrt{4}$
c) $\sqrt{16} \bullet \sqrt{2}$
d) $8\sqrt{2}$

Add:
$(-x^2 + x + 2) + (3x^2 + 4x - 5)$

a) $4x^2 + 5x + 7$
b) $2x^2 + 5x - 3$
c) $2x^2 + 4x - 7$
d) $2x^2 + 3x + 3$

Simplify the expression:

$$\frac{x^2 + 4x + 4}{x^2 + 9x + 14}$$

a) $\frac{1}{5x+10}$
c) $\frac{x^2+2}{(x+2)(x+7)}$
b) $\frac{(x+2)(x+1)}{(x+7)(x+2)}$
d) $\frac{x+2}{x+7}$

Simplify:

$4\sqrt{3} - 2\sqrt{3}$

a) 24
b) $6\sqrt{3}$
c) $2\sqrt{3}$
d) 2

Evaluate $9 + (3 - 1) - 3^2$

a) 8
b) 2
c) 6
d) 0

Find the sum:
$9 + (-12) + 5$

a) 2
b) 26
c) 8
d) 16

Solve:
$9r - 2 = 24 - 4r$

a) $\frac{26}{9}$
b) $\frac{9}{26}$
c) $\frac{1}{2}$
d) 2

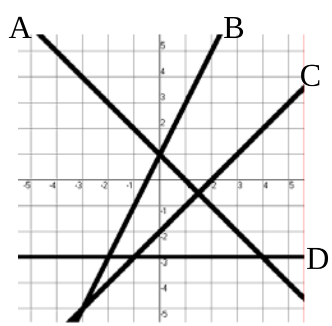
Solve:
$4x - 3 = 13$

a) 4
b) $\frac{13}{4}$
c) 10
d) 16

Find the slope of a line through (-3, 1), (2, 1)

a) $\frac{5}{2}$
b) 0
c) $-\frac{2}{5}$
d) -1

Which line on the graph is $2x + y = 1$?



a) Line A
b) Line B
c) Line C
d) Line D

Write the equation of a line through (-2, -8), (2, 4)

a) $y = 3x + 4$
b) $y = -2x + 8$
c) $y = 3x - 2$
d) $y = 2x + 4$

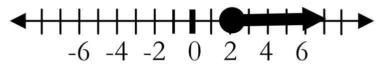
Write the equation in slope-intercept form if $m = 3$ and $b = 2$

a) $y = 3x + 2$
b) $3y = 3x + b$
c) $y = 2x - 3$
d) $y = 3x + 4$

Solve $|2x-3|=5$

a) 4, -1
b) 8, -8
c) 8, -4
d) 1, -1

This graph shows the solution for which equation?



a) $2x < 4$
b) $3x - 5 \geq 4$
c) $-6 \leq -8 + x$
d) $-x > 2$

Solve the linear system:

$2x + 5y = 7$
$7x + y = 8$

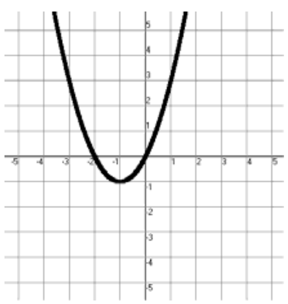
a) (-1, 1)
b) (1, 1)
c) (-2, 7)
d) (7, -8)

Simplify the expression:

$$\frac{a^2}{ab^3} \bullet \frac{b^4}{a^3}$$

a) $\frac{a^8}{a^3b^3}$
b) $\frac{b}{a^2}$
c) $\frac{ab^8}{a^4b^3}$
d) $\frac{b}{a}$

Which function matches this graph?



a) $y = 5x - 7$
b) $y = x^2 + 2x$
c) $y = -2x^2 - 3$
d) $y = x^2 + 8$

Factor this trinomial:

$3x^2 - 8x + 4$

a) $(3x - 2)(x - 2)$
b) $(4x + 2)(-x + 4)$
c) $(3x - 2)(x + 2)$
d) $(2x + 2)(x + 2)$

Simplify the expression:

$$\frac{2x+1}{3x} + \frac{x+5}{3x}$$

a) $\frac{3x+6}{6x}$
b) $\frac{2x+4}{3x}$
c) $\frac{2x^2+11x+5}{9x^2}$
d) $\frac{x+2}{x}$

Solve the equation:

$$\sqrt{x-1} = 5$$

a) $x^2 + 6$
b) 6
c) 26
d) $x - 4$

Algebra Probe Data Collection Procedures
District B Senior High
District C Senior High
October and November 2004

Materials:
1. Student copies of the probes
2. Stopwatch/timer
3. Pencils for students

General Introduction: (*Note: Only do this the first time the class participates in probes.)*
As you all know, your class and other algebra classes at here at District B/District C are working with Iowa State on a research project to learn more about improving algebra teaching and learning. Today I need your help in trying out some of the brief tasks that teachers may be able to use to track student progress. As you may remember, ALL students will complete the tasks, but we will only use your scores in the research project if you and your parent or guardian have both given us permission to do so. Please clear your desk—the only thing you'll need for this activity is a pencil or a pen. (Distribute pencils to any students who need them.)

There are a few things you should know about the tasks, or probes, we will ask you to complete today. First, we will be limiting the amount of time you have to work on the tasks. We EXPECT that you will NOT be able to finish the probes. These tasks are different from classroom tests or quizzes and are not meant to be completely finished. Second, there may be problems on the probes that are difficult or unfamiliar. Please work across each row and try each problem. If you do not know how to answer the question, skip it and go on to the next question. DO NOT spend a great deal of time on any one problem. If you get to the end of the probe and still have time to work, go back to the problems you skipped and try to solve them. Remember, your score on the probe will not hurt your grade in algebra class, but it is important for the research project that you do your best work. Do you have any questions at this point?

Introduction for Subsequent Data Collection Visits:
Today I need your help again to try out some of the short algebra tasks like we did earlier. Just like the last time we did the algebra tasks (which we call probes), there will be time limits for each probe and we EXPECT that you will not be able to complete every problem in the amount of time we give you. What's most important is that you try to find the problems you are able to do and that you give each task your best effort. Are there any questions before we begin?

Additional instructions for days that repeat the same probe sequence previously administered:
Today we will be completing the same set of algebra probes you did last week. The reason we do this is because part of the research is set up to check whether student's scores on the probes are consistent from one week to the next. Is there anyone who was not here the last time we did probes (give date)? If so, the administrator should briefly summarize the directions for the INITIAL administration of each type of probe**.**

After the final probe:
Say, ***That is the end of the tasks for today. In two weeks we will be back in your class to do some more of the probes. Thank you for your help with our research project!***

Directions for the Algebra Basic Skills Probes: **A1, A2**

1. Distribute copies of the first Version A probe to all students in the group FACE DOWN. Ask students to keep the probes face down until they are told to begin.

2. Say to the students:
   This is one type of task we are testing out. The problems on this probe include algebra equations using basic math facts, simplifying expressions by combining like terms, and using the distributive property to simplify expressions. Look at each problem carefully before you answer it.

   Please begin in this corner (demonstrate, pointing to upper left corner) and work down each column, considering each problem. If you do not know how to answer the question, put a small X in the box and continue to the next problem. DO NOT spend a great deal of time on any one problem. When you get to the end of the first column, go on to the second, and then to the second page. If you get to the end of the probe and still have time to work, go back to the problems you marked with an X and try to solve them. On this probe, we are trying to figure out the best amount of time to allow students to work. As you are taking these probes, I will say, "SLASH" at two points. This means that you should draw an obvious slash mark after the problem you are working on. (Demonstrate on board or overhead.) ***Do you have any questions at this point?***

   [DO NOT REPEAT THESE TWO PARAGRAPHS FOR THE SECOND ADMINISTRATION.]

   When I say 'begin,' turn the sheet over and begin answering the problems. Start on the first problem on the left on the top row. Work across and then go the next row. If you can't answer the problem, make an 'X' on it and go on to the next one. Remember to make a slash mark when I say "slash." You will have 4 minutes to work.

3. Set timer for **4** minutes. Say ***Begin*** and start your stopwatch.
4. When timer reads **3** minutes, say ***Slash***.
5. When timer reads **2** minute, say ***Slash***.
6. When timer goes off, say ***Stop. Put your pencils down.***
7. Ask students to pass papers to the back of the room and prepare to repeat for second A version probe. Say, ***Now we will do another probe that is similar to the one you just finished.***

Directions for the Algebra Concepts Probes: **B1, B2**

Directions for Version B Probes

1. Distribute copies of the first Version B probe to all students in the group FACE DOWN. Ask students to keep the probes face down until they are told to begin.

2. Say to the students:
   This is a/another type of task we are testing out. The problems on this probe include translating words into expressions, solving simple equations, interpreting line graphs, and completing function or pattern tables. Look at each problem carefully before you answer it.

   Please begin in this corner (demonstrate, pointing to upper left corner) ***and work across each row, considering each problem. If you do not know how to answer the question, put a small X in the box and continue to the next problem. DO NOT spend a great deal of time on any one problem. If you get to the end of the probe and still have time to work, go back to the problems you marked with an X and try to solve them. On this probe, we are trying to figure out the best amount of time to allow students to work. As you are taking these probes, I will say, "SLASH" at two points. This means that you should draw an obvious slash mark after the problem you are working on.*** (Demonstrate on board or overhead.) ***Do you have any questions at this point?***

   [DO NOT REPEAT THIS PARAGRAPH FOR THE SECOND ADMINISTRATION.]

   When I say 'begin,' turn the sheet over and begin answering the problems. Start on the first problem on the left on the top row. Work across and then go the next row. If you can't answer the problem, make an 'X' on it and go on to the next one. Remember to make a slash mark when I say "slash." You will have 6 minutes to work.

3. Set timer for **6** minutes. Say ***Begin*** and start your stopwatch.
4. When timer reads **2** minutes, say ***Slash***.
5. When timer reads **1** minute, say ***Slash***.
6. When timer goes off, say ***Stop. Put your pencils down.***
7. Ask students to pass papers to the back of the room and prepare to repeat for second B version probe. Say, ***Now we will do another probe that is similar to the one you just finished.***

Directions for the Content Analysis Probes: **C1, C2**

1. Distribute copies of the Version C probe to all students in the group FACE DOWN. Ask students to keep the probes face down until they are told to begin.

2. Say to the students:
This is a third type of task we are testing out. The problems on this probe represent the different types of problems that you are learning in your textbook. In general, you will probably find that the problems at the beginning are easier and those on the second page are more challenging. Look at each problem carefully before you answer it. The problems on this probe may seem more difficult than those on the probes you've already completed. If you would like to skip around as you answer the problems on this probe, you can do so; I will not ask you to make slashes during this probe.

3. ***When I say 'begin,' turn the sheet over and begin answering the problems. Start on the first problem on the left on the top row. Work across and then go the next row. If you can't answer the problem, make an 'X' on it and go on to the next one. Remember to make a slash mark when I say "slash." You will have 7 minutes to work.***

4. Set timer for **7** minutes. Say ***Begin*** and start your stopwatch.

5. When timer goes off, say ***Stop. Put your pencils down.***
6. Ask students to pass papers to the back of the room.

<u>Directions for Reform-Oriented Probes</u>: **D1, D2**

1. Distribute copies of the first Version D probe to all students in the group FACE DOWN. Ask students to keep the probes face down until they are told to begin.

2. Put the Sample Page on the overhead and say to the students:
   This page shows an example page from [the first/another] type of task we are testing out. At the top of the page, you will see a row of graphs. On other pages, the top row may have tables or equations in it. Below this top row (**point*****), you will see a set of boxes. As you move to the second row, your task is to match each item to one of the boxes in the first row. Let's look at this sample page together. The first equation in the second line says y = x + 4. Which graph does this equation match?***
   Pause and wait for students to identify graph C. If they do not, say:
   In this equation, if x is 0, what would y be? [4] Do you see a graph that has the point (0, 4) on it? Yes, that's right; graph C has the point (0, 4) as part of the line. So this answer would be C (demonstrate how to write answer in the blank).
   Good! Now let's look at another equation in this row. (Point to the equation y = 0 in the fourth box)***. This equation says y = 0. Do you see a graph that shows that no matter what value x has, y will always be 0?*** (pause, wait for students to say graph B). ***Yes, that's right, graph B matches the equation y = 0, so we'll write a B on the line next to this equation*** (model writing answer on blank). ***Now let's look at the tables in the 3rd row. If we look at the first table, can you figure out which graph goes with this set of values for x and y***? (Pause, wait for students to say Graph D). ***Yes, that's right. Graph D includes the pointes (2, -1), (1, 0), and (0, 1). So for this blank, we would write a D in the blank. This final section in this probe has several different story situations. Let's read the first one together***. (Read Bill story aloud). ***Which graph would match this story scenario***? (Pause. Wait for students to identify Graph C)***. Yes, that's right. Graph C shows the relationship. As you work on the story scenarios, it is important to remember that the story may apply to just a portion of the graph or table. For example, in the Bill story, only the part of the graph where both x and y are positive fits with the story. As you work on this part, you can ask to have a word or story situation read aloud to you if that would be helpful. Just raise your hand and I*** (or teacher, if available) ***will read it for you. Do you have any questions about how to do this type of probe?*** [DO NOT REPEAT THIS PARAGRAPH FOR THE SECOND ADMINISTRATION. GO DIRECTLY TO THE PARAGRAPH BELOW.]

   [Please put your name, date, and period on the back page.] This is an algebra probe that requires you to match each item to the four lettered graphs, equations, or tables at the top of each page. You will have 7 minutes to work, so be sure not to spend too much time on any one problem. Please do your best work. When I say 'begin,' turn the probe over and begin answering the problems.

3. Set timer for **7** minutes. Say ***Begin*** and start your stopwatch.
4. When timer goes off, say ***Stop. Please put your pencils down.***
5. Ask students to pass papers to the back of the room and prepare to repeat for second D version probe. Say, ***Now we will do another probe that is similar to the one you just finished.***

Directions for Content Analysis-Multiple Choice Probes: **E1, E2**

1. Distribute copies of the first Version E probe to all students in the group FACE DOWN. Ask students to keep the probes face down until they are told to begin. ***Please put your name, today's date, and the period on the back of this sheet.***

2. Say to students,
This is [the first/another] type of task we are testing out. The problems on this probe represent the different types of problems that you are learning in your textbook. Each question has four multiple choice options that you can choose from for your answer. In general, you will probably find that the problems at the beginning are easier and those on the second page are more challenging. Look at each problem carefully before you answer it.

   On this probe, you get points both for the answer and for showing the work that you did. Even if you don't get the final answer correct, you can still earn points for showing your work. If you can do the problem in your head, you can get full credit for doing that. HOWEVER, you need to be careful about making wild guesses on the multiple choice questions. If you choose an incorrect answer, you will lose a point, so it is better to skip a problem than to make a wild guess. Do you have any questions about how to do this probe?
[DO NOT REPEAT THIS PARAGRAPH FOR THE SECOND ADMINISTRATION.]

   [Please check to see that your name, date, and period are on the back of your paper.] We are going to do an algebra probe uses a multiple choice format. When I say 'begin,' turn the sheet over and begin answering the problems. Remember that you can earn points on each problem by showing your work, even if you aren't able to complete the entire problem. You should NOT make wild guesses and just circle any answer when you don't know how to do a problem. Please do your best work. You will have 7 minutes to work.

3. Set timer for **7** minutes. Say ***Begin*** and start your stopwatch.
4. When timer goes off, say ***Stop. Put your pencils down.***
5. Ask students to pass papers to the back of the room and prepare to repeat for second E version probe. [Say, ***Now we will do another probe that is similar to the one you just finished.]***

Teacher ____________________

QuickTime™ and a
TIFF (LZW) decompressor
are needed to see this picture.

Project AAIMS:
Algebra Assessment and Instruction:
Meeting Standards

Directions:	Below is a list of the students you teach. Please rate each student's proficiency in algebra in comparison to others in the student's grade. A rating of "1" indicates a low level of proficiency compared to others in the same grade, "3" indicates average proficiency and "5" indicates a high level of proficiency.

Student	Algebra Proficiency				
	Low		Average		High
	1	2	3	4	5
	1	2	3	4	5
	1	2	3	4	5
	1	2	3	4	5
	1	2	3	4	5
	1	2	3	4	5
	1	2	3	4	5
	1	2	3	4	5
	1	2	3	4	5
	1	2	3	4	5
	1	2	3	4	5
	1	2	3	4	5
	1	2	3	4	5
	1	2	3	4	5
	1	2	3	4	5