PROJECT AAIMS: ALGEBRA ASSESSMENT AND INSTRUCTION – MEETING STANDARDS

Reliability and Criterion Validity of Four Revised Algebra Measures in Districts B and C

Technical Report #10

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 Four Revised Algebra Measures in Districts B and C

Technical Report 10

Abstract

This technical report summarizes the results of a study in which we examined the technical adequacy of four revised measures for algebra progress monitoring. The measures investigated included a *Basic Skills* probe, an *Algebra Foundations* probe, a *Translations* probe, and a *Content Analysis-Multiple Choice* probe. Revisions to the measures included the addition of a sample page prior to the initial administration of each type of probe and changes to the design templates used to create the *Content Analysis-Multiple Choice* measure. Seventy-eight students (6 of whom were receiving special education services) completed two forms of a *Basic Skills* measure, two forms of an *Algebra Foundations* 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 then 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, as well as criterion validity, for both single probe scores and aggregated scores (computed by averaging two individual scores).

We found that three of the four measures produced effective distributions of student scores, with no signs of floor or ceiling effects. On the *Translations* probe, students produced nearly half 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 was higher than results obtained in previous studies, with correlations for most measures (except *Translations*) in the .6 to .8 range. Aggregating scores from two probes produced reliability estimates in the .7 to .8 range for all measures except the *Translations* measure.

Criterion validity was examined by correlating students' scores on the probes with other indicators of proficiency in algebra, including grades in algebra, teacher ratings, scores on the district's achievment test, and scores on a standardized test of algebra aptitude. Correlation coefficients were higher than those obtained in earlier studies (Foegen & Olson, 2005; Foegen, Olson, & Perkmen, 2005). Correlation coefficients for single probes were generally in the low to moderate range (.3 to .5); small increments in the coefficients were obtained when aggregared scores were used. The strongest relations were obtained for the *Basic Skills* and the *Content Analysis-Multiple Choice* measures, followed by the *Algebra Foundations* measure. Students' scores on the probes are most related to their teachers' ratings, their grades in algebra, and their scores on the algebra aptitude test. 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, we reported the reliability and criterion validity of three measures developed as potential indicators of student proficiency in algebra. In Technical Report 6, 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 Technical Report 7, we reported the reliability and criterion validity of all five potential algebra measures in a study conducted in two of the districts participating in Project AAIMS. In this report, we describe a follow-up study in the same two districts with four of the potential measures of algebra proficiency. We also describe the revisions to the content and administration procedures that we incorporated based on the results of the study reported in Technical Report 7.

Method

The study described in this report was conducted from March to May 2005 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. Approximately 15% of the student population (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 approximately 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 a six-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

Sixty-two students in District B and 16 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	62[a]	38	20	3	0
Gender					
Male	29	18	10	1	0
Female	33	20	10	2	0
Ethnicity					
White	51	31	18	2	0
Black	9	6	1	1	0
Hispanic	2	1	1	0	0
Native Am.	0	0	0		
Lunch					
Free/Red	26	14	9	2	0
Disability					
IEP	6	0	5	1	0

[a] The grade level totals do not add to 62 because grade data were missing for 1 participant.

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

	Total	Grade 9	Grade 10	Grade 11	Grade 12
N	16	11	5	0	0
Gender					
Male	7	6	1	0	0
Female	9	5	4	0	0
Ethnicity					
White	13	10	3	0	0
Black	1	1	0	0	0
Hispanic	2	0	2	0	0
Native Am.	0				
Lunch					
Free/Red	5	2	3	0	0
Disability					
IEP	0	0	0	0	0

As the data in Tables 1 and 2 indicate, many of the participants (an average of 82%) were white and an average of 63% were in ninth grade, the traditional grade in which students in these districts complete algebra. 42 and 31 percent participated in federal free and reduced lunch programs in Districts B and C, respectively and 10% of the participating students in District B were students with disabilities who were receiving special education services. In District B, 13 students were enrolled in Algebra 1, 5 in Algebra 1A, and 44 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 each of two single courses. 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 five of the six students with disabilities in District B participating in the project is provided in Table 3. Data were not reported for one of the students.

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

Characteristic	Quantification
Disability category	5 Entitled Individual (EI)
% time in general education	Range = 97 –100%; Mean = 99% 4 of the 5 students spend 100% of their instructional time in general education
# of students with math goals	1
# of students receiving math instruction in general education classes	5
# of students receiving math instruction in a special education setting	0
# of students with goal code F2C: *Comprehension*	3
# of students with goal code F3A: *Applied math*	0
# of students with goal code F3C*:* *Computation*	1
# of students with goal code F4M: *Mechanics of writing: punctuation, grammar, spelling*	4

Students with disabilities earned a mean GPA for the 2004-05 academic year of 2.55 (range 1.60 – 3.18). In algebra, students with disabilities earned mean grades of 2.06 [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 in District B with disabilities obtained national percentile rank scores of 32 and 50 in Concepts/Problem Solving, and Computation, respectively. They demonstrated greater deficits in reading, with mean percentile ranks of 27 for the Reading Total score.

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 of the groups of measures is described below.

Algebra Progress Monitoring Measures. Four algebra measures were examined in this study; sample copies of each are provided in the Appendix. The following paragraphs summarize the characteristics of each of the four types of algebra measures. One of the measures used in previous studies (*Content Analysis-Constructed Response*) was dropped for this study because of the low levels of inter-scorer reliability we had obtained in previous studies and the inordinate amount of time required to score the probe. While the measure had promising data for reliability and validity, we decided the time requirements for scoring were so great that they prevent the measure from being implemented in a typical school setting without the resources available through an outside source, such as Project AAIMS. The final paragraph for each probe outlines the changes that were made to the probe from the fall study (Technical Report 7) to this study.

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 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.

For this study, we decreased the proportion of problems involving solving simple equations using basic facts (e.g., $5 + b = 12$; $b =$ ___) from 50% of the problems on the earlier version of the *Basic Skills* probe to 33%. We opted to make this change because earlier results (Technical Reports 2 and 7; Foegen & Lind, Foegen, Olson, & Perkmen, 2005) had revealed stronger correlations with the computation subtest of the districts' standardized achievement measure than with the problem solving subtest. We were concerned that having half of the problems of this type emphasized computation skills to a greater degree than we had intended in relation to the other types of problems on the probe.

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 variables and expressions; (2) manipulating expressions involving integers, exponents, and order of operations; (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 with 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 require two responses, so 50 total points were possible on this probe.

The only changes we made to the *Algebra Foundations* measure for this study were slight revisions to make the format of the items consistent (i.e., using *Solve* or *Evaluate* as prompts for different types of problems). We also changed one of the four coordinate graphing problems from a task requiring students to identify the *y*-intercept and the slope of a line to a problem in which students identified the coordinates for two points on the graph.

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 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.

For this study, we re-formatted the problems on each page of the measure to change which type of representation was used at the top of the page as the prompt to which students were to match the subsequent representation formats. For the earlier studies, the first page used graphs as a prompt, followed by data tables on the second page, and finally equations on the third. Drawing on feedback from participating teachers, we used equations as the prompt for the first page, data tables on the second page, and graphs on the third page.

The *Translations* probe was created in response to feedback from the Project AAIMS Advisory Committee during the 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 (Probe C)* probe examined in the initial study. For this measure, we revised to 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 in the study reported in Technical Report 2, Foegen & Lind, 2004), 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.

For this study, we revised the *Content Analysis-Multiple Choice* measure by refining the item template used to create each probe. In our earlier studies (reported in Technical Reports 6 and 7, Foegen & Olson, 2005, Foegen, Olson, & Perkmen, 2005), we identified a low rate of alternate form reliability. We were concerned that this might be an artifact of the process used to create problems for each probe. In the earlier versions of this measure, one to three key concepts were identified for each chapter of the text and we sampled from these key concepts to create problems. In the event that a chapter was represented on the probe by only one problem, but contained more than one key concept, the problem selected for this chapter on Probe 1 might be entirely different than the problem created for Probe 2. To address this inconsistency across forms of the measures, we met with participating teachers in January and February 2005 and worked with them to streamline our probe templates. The teachers helped us to identify (and in some cases merge) key ideas that corresponded to each question on the probe. As a result, we were able to create multiple forms of the probe that we believed would be more comparable and, as a result, produce more acceptable levels of alternate form reliability.

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.

<u>Criterion Measures</u>. 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, classroom performance 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-point 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.

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). 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, while students in District C complete the ITED in March. 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 7th grade students for successful study of algebra in 8th 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

Data for the study were collected from March to May 2005. In each class, students completed the IAAT first, then one to two weeks later began a series of four weeks of probe data collection. During this time, participating teachers completed the teacher rating forms and assisted project staff in obtaining data from students' school records.

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 four different types of algebra probes into two groups. One group consisted of two parallel forms of the *Basic Skills* probe and two forms of the *Translations* 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. Students completed the tasks in the same order both weeks. Copies of the standardized directions used for each administration session are provided in the Appendix. Table 4 depicts the order in which the probes were administered during each of the four testing sessions for the algebra probes.

Table 4. Administration Schedule for Probe Forms by Period

Session	S-2 (1A)	K-3 (1B)	E-4 (1B)	S-4 (1A)	K-4 (1A)	B-2 (1)	B-4 (1)
1 and 2	A21	B11	D11	E21	A22	B12	D12
	A22	B12	D12	E22	A21	B11	D11
	D11	E21	A21	B11	D12	E22	A22
	D12*	E22	A22	B12	D11	E21	A21
3 and 4	B11	A21	E21	D11	B12	A22	E22
	B12	A22	E22	D12	B11	A21	E21
	E21	D11	B11	A21	E22	D12	B12
	E22	D12	B12	A22	E21	D11	B11

A21, A22 = *Basic Skills* probes 1 and 2
B11, B12 = *Algebra Foundations* probes 1 and 2
D11, D12 = *Translations* probes 1 and 2
E21, E22 = *Content Analysis-Multiple Choice* probes 1 and 2
* Probe D12 was not administered in this class in Session 1 because sufficient time was not available due to a school assembly.

Scoring

The scoring for the *Basic Skills* and *Algebra Concepts* probes was completed by counting the number of problems completed correctly. We used the results from the study described in Technical Report 7 to select five minute durations as the time limits for these probes. For each of these probes, the scoring process included determining the number of problems completed in each of the three durations. The scoring for the remaining two probes, *Translations* and *Content Analysis-Multiple Choice*, was identical to the procedures described in Technical Report 7. 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 a portion 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 seven class periods across the four administration periods. Each form of the probes was rescored for three class periods drawn from the 14 total administrations (21%). 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	83	87.5 - 100	97.6%
Algebra Foundations	66	83.9 - 100	94.4%
Translations	66	87.5 - 100	99.4%
Content Analysis-Multiple Choice	75	66.7 - 100	94.6%

The *Basic Skills* and *Translations* probes were scored with the highest level of accuracy. The *Algebra Foundations* and *Content Analysis-Multiple Choice* had slightly lower, but still acceptable, accuracy levels (just under 95%). These levels are somewhat lower than those reported in Technical Report 6 (Foegen, Olson, & Perkmen, 2005), where we obtained interscorer agreement levels of 97.7% and 97.6% for the *Algebra Foundations* and *Content Analysis-Multiple Choice* measures, respectively. We are uncertain as to why the interscorer reliability levels dropped off somewhat in the present study.

<u>Descriptive Data on Score Ranges and Distributions</u>

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 43 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	Range	Mean	Standard Deviation
Basic Skills Form 1	1	69	5 min.	1 – 51	23.04	9.40
	2	69	5 min.	3 – 57	26.54	10.28
Basic Skills Form 2	1	68	5 min.	3 – 49	22.06	10.13
	2	70	5 min.	1 – 56	26.26	10.68
Algebra Foundations Form 1	1	65	5 min.	1 – 37	23.00	6.80
	2	65	5 min.	4 – 45	25.83	7.61
Algebra Foundations Form 2	1	63	5 min.	9 – 39	24.83	7.12
	2	64	5 min.	10 – 44	27.33	8.28
Translations Form 1	1	68	Correct	2 – 34	15.47	6.75
		68	Incorrect	0 – 37	10.41	8.72
	2	69	Correct	4 – 29	16.10	6.55
		69	Incorrect	0 – 36	12.74	10.48
Translations Form 2	1	57	Correct	2 – 32	15.39	6.16
		57	Incorrect	0 – 31	8.05	7.27
	2	70	Correct	3 – 30	17.23	6.23
		70	Incorrect	0 – 32	8.66	8.26
Content Analysis-Multiple Choice Form 1	1	65	Correct	6 -37	19.86	6.62
		65	Incorrect	0 – 13	2.08	3.07
	2	63	Correct	3 – 38	21.95	8.80
		63	Incorrect	0 – 15	2.24	3.19
Content Analysis-Multiple Choice Form 2	1	67	Correct	0 - 35	20.06	6.41
		67	Incorrect	0 – 11	1.96	2.79
	2	69	Correct	5 – 38	20.72	7.41
		69	Incorrect	0 - 12	2.43	3.12

For both the *Basic Skills* and *Algebra Foundations* measures, there was no evidence of floor or ceiling effects. The mean scores on each probe increased three to four points from the first administration to the second. The standard deviations were substantial (one-third to one-half the magnitude of the mean), suggesting that the measures would be effective in spreading out students on the basis of proficiency on these tasks.

On the *Translations* measure, we obtained results similar to those of previous studies. The mean number of problems Incorrect was large in relation to the number of problems Correct. In general, the mean number of Incorrect responses was half as large (or more) than the number of Correct responses. In addition, the standard deviations for the Incorrect responses exceeded those for the Correct responses. In previous research, we attributed these results to a high rate of student guessing, perhaps in part due to the fact that students were not explicitly instructed NOT to make wild guesses when responding to the *Translations* probes. In this study, students were specifically instructed that making random guesses would lower their score on the probe. However, this instruction does not appear to have had a substantial effect in lowering the rate of guessing.

The *Content Analysis-Multiple Choice* measure results did not reveal any concerns regarding floor or ceiling effects. The number of Incorrect responses was low relative to the number of Correct responses (often only about one-tenth), indicating limited problems with guessing and/or random responses. The standard deviations for the number of Correct responses were about one-third the number of Correct responses, again supporting the potential for this measure to discriminate between high and low performing students.

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 a 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			5 min.				5 min.	
Session 1			.80		Form 1		.88	
Session 2			.89		Form 2		.71	
Algebra Foundations			5 min.				5 min.	
Session 1			.59		Form 1		.64	
Session 2			.68		Form 2		.71	
Translations		Corr	C – I	C -1/3		Corr	C – I	C -1/3
Session 1		.52	.52	.50	Form 1	.56	.46	.52
Session 2		.60	.65	.66	Form 2	.58	.46	.52
Content Analysis-Multiple Choice		Corr	C – I	C -1/3		Corr	C – I	C -1/3
Session 1		.59	.66	.61	Form 1	.72	.80	.76
Session 2		.75	.76	.75	Form 2	.70	.74	.71

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

The reliability of single forms of the *Basic Skills* probe was quite strong, ranging from .71 to .89. Alternate form reliability was higher than test-retest reliability. For the *Algebra Foundations* measure, reliability ranged from .59 to .71, with test-retest reliability higher than alternate form reliability. These reliability estimates are well below the .80 standard often used as a benchmark for evaluating reliability. It may be necessary to administer multiple forms in order to get a more stable estimate of student performance on this measure.

The *Translations* measure reliability estimates ranged from .46 to .66. We did not obtain consistent improvements when the corrected scoring procedures were applied. In general, alternate form reliability was higher than test-retest reliability, but both fell far short of acceptable levels. This result, which has now been replicated in three studies (the present study as well as Technical Report 6 (Foegen & Olson, 2004) and Technical Report 7 (Foegen, Olson, & Perkmen, 2005), raises strong concerns about the viability of the *Translations* measure for future use in the participating districts. We discuss this finding further in the Discussion section.

The reliability of the *Content Analysis-Multiple Choice* measure ranged from .59 to .80. Test-retest reliability was stronger than alternate form reliability. The Correct – Incorrect scoring procedure produced the strongest coefficients in each set of analyses. The lowest coefficients were obtained for alternate form reliability in Session 1. In Session 2, these coefficients increased by .10 to .16. This finding suggests that as students become more familiar with this task, their scores are likely to become more consistent. The data for this measure were much stronger than those obtained for the earlier version of the *Content Analysis-Multiple Choice* probe reported in Technical Report 7 (Foegen, Olson, & Perkmen, 2005). In the earlier study, reliability estimates for single probes of this type ranged from .47 to .61. The increases in reliability are likely a result of the revisions made to the probe template intended to increase consistency across multiple forms.

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

Table 8. Reliability for Aggregated Probes

Probe	Alternate Form Reliability	Test-Retest Reliability
Basic Skills 5 min.	.88	.85
Algebra Foundations 5 min.	.73	.76
Translations		
Correct	.62	.66
C - I	.60	.72
C - 1/3	.69	.64
Content Analysis-Multiple Choice		
Correct	.72	.76
C - I	.74	.80
C - 1/3	.72	.76

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

The reliability of aggregating two scores was quite strong for the *Basic Skills* measure (.85 to .88). Aggregation also produced considerable improvements in the *Algebra Foundations* reliability levels, although further improvements in reliability on this measure should be pursued. A mixed pattern of results was obtained when scores on the *Translations* measure were aggregated. In some cases, the reliability level improved, while in others, it decreased. Aggregating scores on this measure did not produce substantial improvements in reliability for this measure. On the *Content Analysis-Multiple Choice* probes, aggregating scores resulted in increases in alternate form reliability over those obtained in Session 1, but decreases in comparison to single form reliability for Session 2. The test-retest reliability of aggregated scores was the same as or slightly higher than the levels obtained for single probes. In each case, the Correct – Incorrect scoring procedure produced the highest reliability levels.

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 overall GPA; teachers' evaluations of student proficiency; scores from standardized tests in mathematics administered by the district (ITED); 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 are presented in Tables 9 to 11, with results included for 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 9 includes correlations of single algebra measures with grade-based criterion measures and teacher ratings. Table 10 includes correlations between single algebra measures and ITED subtest scores. Table 11 includes correlations between single algebra measures and IAAT subtest and total test scores.

Table 9. Criterion Validity Results for Single Probes: Mean Correlation Coefficients and Ranges for Grade-based Measures and Teacher Ratings

		Overall GPA	Grade in Algebra	Teacher Rating
Basic Skills	5 min.	.35 (1 NS; .28 - .41)	.53 (.45 - .57)	.56 (.53 - .59)
Algebra Foundations	5 min.	NS (3 NS; .30)	.38 (.29 - .43)	.44 (1 NS, .38 - .48)
Translations	Correct	NS (3 NS; .27)	.31 (1 NS, .27 - .35)	NS
	C – I	NS (3 NS; .34)	.35 (2 NS, .34 - .36)	NS
	1/3	NS (3 NS; .35)	.33 (1 NS, .26 - .36)	NS
Content Analysis-Multiple Choice	Correct	.35 (2 NS, .33 - .37)	.45 (.34 - .57)	.41 (1 NS, .33 - .50)
	C – I	.38 (1 NS, .30 - .43)	.47 (.34 - .59)	.44 (1 NS, .38 - .53)
	1/3	.35 (1 NS, .27 - .40)	.46 (.35 - .57)	.43 (1 NS, .35 - .52)

Correlations between the four types of algebra probes and grade-based measures and teacher ratings were in the low to moderate range. Overall GPA showed limited relations to students' scores on the probes. For two of the measures (*Algebra Foundations* and *Translations*), the relations were non-significant. For the remaining two measures, correlations were in the low range (.35 to .38). Stronger relations were obtained with students' algebra grades. The strongest relation was with the *Basic Skills* measure (.53), followed by the *Content Analysis-Multiple Choice* measure (.47 for Corrrect – Incorrect scoring). Relations with the *Algebra Foundations* and *Translations* measures were in the low range (.35 to .38). Three of the four measures showed moderate relations with teacher ratings. Again, the *Basic Skills* measure was strongest (.56), followed by the *Algebra Foundations* and *Content Analysis-Multiple Choice* measures (both .44). The *Translations* measure did not demonstrate a statistically significant relation to teacher ratings.

Table 10. Criterion Validity Results for Single Probes: Mean Correlation Coefficients and Ranges for ITED Scores

		Con/Prob	Comp	Reading Total
Basic Skills	5 min.	NS (3 NS; .27)	.39 (.34 - .44)	NS
Algebra Foundations	5 min.	NS (3 NS; .26)	NS (3 NS; .30)	NS
Translations	Correct	NS	NS	NS
	C – I	NS (3 NS; .27)	.31 (2 NS, .29 - .33)	NS
	1/3	NS	NS (3 NS; .30)	NS
Content Analysis-Multiple Choice	Correct	NS (3 NS; .29)	.26 (2 NS, .26 - .26)	NS
	C – I	NS	NS (3 NS; .27)	NS
	1/3	NS (3 NS; .26)	.25 (2 NS; .24 - .26)	NS

Students' scores on the algebra probes had limited to no relation to their performance on the district's achievement test, the Iowa Test of Education Development (ITED). None of the measures demonstrated statistically significant relations with the Concepts/Problems subtest of the ITED. The Computation subtest demonstrated correlations in the low range with the *Basic Skills* measure (.39), the Correct – Incorrect scoring of the *Translations* measure (.31) and the Correct and Correct – 1/3 Incorrect scoring for the *Content Analysis-Multiple Choice* measure (.26 and .25, respectively). None of the measures were related to students' performance on the Reading Total scale of the ITED, suggesting that students' reading ability was not associated with their performance on the algebra measures.

These results are not entirely unexpected. The content of the ITED does not emphasize algebra proficiency. Instead, the test gauges general mathematics proficiency (i.e., basic skills), rather than proficiency in the advanced content (algebra, geometry) often taught in high school mathematics.

Table 11 summarizes the correlations between the four types of algebra probes and the Iowa Algebra Aptitude Test. Results are reported for the Total scale score, as well as for each of the subtests. Part A of the IAAT measures Interpreting Mathematical Information, while Part B assesses Translating to Symbols. Part C focuses on Finding Relationships, while Part D involves Using Symbols. The *Basic Skills* and *Algebra Foundations* measures had the strongest relations with the IAAT Total Scale (.55 and .52, respectively). For the *Content Analysis-Multiple Choice* measure, the Correct – Incorrect coefficient (.42) was slightly larger than those for Correct – 1/3 Incorrect and Correct alone. The *Translations* measure was least related (.30 to .32) to the IAAT Total Scale. Among the subtests, the strongest relations with the probe were identified for Parts

Table 11. Criterion Validity Results for Single Probes: Mean Correlation Coefficients and Ranges for IAAT

	IAAT Scores				
	Total	Part A	Part B	Part C	Part D
Basic Skills					
5 min.	.55 (.48 - .60)	.31 (.25 - .37)	.43 (.38 - .48)	.48 (.45 - .52)	.46 (.41 - .48)
Algebra Foundations					
5 min.	.52 (.40 - .66)	.36 (.27 - .48)	.35 (1 NS, .31 - .42)	.41 (.30 - .54)	.44 (.41 - .53)
Translations					
Correct	.30 (2 NS, .25 - .34)	NS (3 NS; .25)	NS	NS (3 NS; .25)	.31 (.29 - .38)
C – I	.30 (.27 - .36)	NS	NS (3 NS; .25)	.31 (1 NS, .26 - .35)	.36 (.32 - .41)
1/3	.32 (.27 - .42)	NS (3 NS; .27)	NS (3 NS; .30)	.30 (1 NS, .26 - .33)	.38 (.33 - .48)
Content Analysis-Multiple Choice					
Correct	.40 (.30 - .45)	NS (3 NS; .35)	.27 (1 NS, .25 - .28)	.37 (2 NS, .36 - .37)	.43 (.31 - .50)
C – I	.42 (.29 - .48)	NS (3 NS; .35)	.27 (.25 - .28)	.37 (1 NS, .33 - .40)	.49 (.34 - .59)
1/3	.41 (.30 - .47)	NS (3 NS; .36)	.27 (.26 - .28)	.35 (1 NS, .28 - .39)	.46 (.32 - .54)

C and D. Again, the *Basic Skills, Algebra Foundations,* and *Content Analysis-Multiple Choice* measures were most correlated to the IAAT subtest scores. The magnitude of these correlations ranged from .27 to .49.

<u>Criterion Validity for Aggregated Probe Scores</u>

In Tables 12 to 14, we report the criterion validity coefficients using 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).

Table 12. Criterion Validity Results for Aggregated Probes: Mean Correlation Coefficients and Ranges for Grade-based Measures and Teacher Ratings

		Overall GPA	Grade in Algebra	Teacher Rating
Basic Skills	5 min.	.31 (.25 - .40)	.53 (.49 - .59)	.58 (.56 - .60)
Algebra Foundations	5 min.	.29 (2 NS, .27 - .30)	.43 (.36 - .47)	.39 (.29 - .48)
Translations	Correct	NS (3 NS; .26)	.34 (1 NS, .31 - .36)	NS
	C – I	.32 (2 NS, .31 - .32)	.31 (.26 - .36)	NS (3 NS; .25)
	1/3	.31 (2NS, .30 - .31)	.33 (.26 - .35)	NS
Content Analysis-Multiple Choice	Correct	.31 (.26 - .38)	.49 (.36 - .59)	.40 (.29 - .50)
	C – I	.37 (.29 - .44)	.50 (.38 - .61)	.41 (.31 - .51)
	1/3	.31 (.26 - .39)	.49 (.36 – 59)	.40 (.29 - .50)

The use of aggregated scores resulted in the greatest improvements for the *Content Analysis-Multiple Choice* measure, for which every coefficient increased over the results obtained when single scores were used. A similar pattern was obtained for the *Basic Skills* measure, which either increased or remained constant. The results were less beneficial for the *Algebra Foundations*, which showed an increase in the correlation for algebra grade, but decreases in the correlations for overall GPA and teacher ratings.

Table 13. Criterion Validity Results for Aggregated Probes: Mean Correlation Coefficients and Ranges for ITED Scores

		ITED Scores		
		Con/Prob	Comp	Reading Total
Basic Skills	5 min.	NS (3 NS; .24)	.40 (.34 - .41)	NS
Algebra Foundations	5 min.	NS	.29 (2 NS, .27 - .31)	NS
Translations	Correct	NS	NS	NS
	C – I	NS (3 NS; .24)	.29 (2 NS, .24 - .34)	NS (3 NS; .23)
	1/3	NS	NS (3 NS; .26)	NS
Content Analysis-Multiple Choice	Correct	NS (3 NS; .26)	.27 (1 NS; .25 - .29)	NS
	C – I	NS	.26 (2 NS, .25 - .26)	NS
	1/3	NS (3 NS; .26)	.27 (1 NS; .25 - .29)	NS

Table 14. Criterion Validity Results for Aggregated Probes: Mean Correlation Coefficients and Ranges for IAAT Scores

	IAAT Scores									
	Total		Part A		Part B		Part C		Part D	
Basic Skills										
5 min.	.56	(.53 - .60)	.33	(.27 - .37)	.43	(.41 - .45)	.50	(.49 - .50)	.46	(.44 - .51)
Algebra Foundations										
5 min.	.57	(.49 - .63)	.40	(.35 - .46)	.36	(.31 - .41)	.46	(.36 - .55)	.48	(.45 - .51)
Translations										
Correct	.31	(2 NS, .30 - .32)	NS	(3 NS; .24)	NS		NS	(3 NS; .26)	.34	(.31 - .38)
C – I	.35	(.32 - .38)	NS		NS		.34	(.31 - .36)	.41	(.39 - .42)
1/3	.36	(.30 - .42)	NS	(3 NS; .26)	NS	(3 NS; .25)	.33	(.31 - .36)	.44	(.39 - .48)
Content Analysis-Multiple Choice										
Correct	.41	(.37 - .47)	.29	(2 NS, .26 - .31)	.28	(.26 - .29)	.33	(.32 - .36)	.45	(.39 - .51)
C – I	.42	(.35 - .50)	.28	(2 NS, .25 - .30)	.27	(.25 - .28)	.35	(.31 - .40)	.50	(.43 - .59)
1/3	.41	(.37 - .47)	.29	(2 NS, .26 - .31)	.28	(.26 - .39)	.34	(.32 - .36)	.45	(.39 - .51)

The results for the ITED scores presented in Table 13. Not surprisingly, given the limited attention to algebra on this measure, aggregating two probe scores had no effect on the relation between students' scores on the probes and their performance on the ITED. As presented in Table 14, the use of aggregated scores generally produced similar results or increased in the correlations in comparison to the coefficients obtained when single probe scores were used. This effect was most pronounced for the *Algebra Foundations* measure. With respect to the alternative scoring procedures, the data support the use of the Correct – Incorrect procedure for the *Content Analysis-Multiple Choice* measure. No clear pattern of results favoring either correction procedure was obtained for the *Translations* measure.

<u>Results for Students with Disabilities</u>

Because a primary focus of Project AAIMS is improving the algebra outcomes for students with disabilities and those at risk for low mathematics achievement, we were also curious about whether differential results might be obtained if students with disabilities were considered as a separate group. In this study, only six students with IEPs were included in the sample, so this presented potential problems due to the small size of the group. We conducted ***exploratory*** correlation analyses to examine the criterion validity of the four different measures. To do this, we correlated the single and aggregated scores for the students with disabilities with their algebra grades, the teachers' ratings of their proficiency in algebra, and the IAAT Total Scale score. We also looked at correlations between students' scores on the measures and ITED results for Concepts/Problems and Computation*.* Table 15 presents the mean correlations between the probe scores obtained by students with disabilities and their scores on other indicators of algebra proficiency. The first column presents the results for individual or single probes, while the second column presents the results obtained for aggregated scores based on the average of two probes. Readers should note that all correlation coefficients were included in the analyses, regardless of statistical significance levels because the small sample size precluded limiting our analyses to statistically significant results. While we might have turned to non-parametric tests, we viewed these analyses as largely exploratory and therefore opted to use the same metric as used for the analyses of the full sample.

The results suggest that for students with disabilities, the *Basic Skills* and *Translations* measures may be best for assessing student proficiency in algebra. The correlations we obtained, though they must be interpreted cautiously, were considerably stronger than those obtained for the sample as a whole. This is especially surprising for the *Translations* measure, about which we had great concerns for the entire sample. It may be that for students with disabilities, this more conceptual means of assessment allows students who understand basic algebra concepts to demonstrate this knowledge in a way that is not possible on the other probes, which emphasize procedural understanding. It is also interesting to note that the correction procedures for the *Translations* and the *Content Analysis-Multiple Choice* measures seemed to be less useful for the students with disabilities. It may be that the students either respond to fewer problems or produce fewer errors so that correcting for random errors has less of an influence on the final score.

Again, readers are reminded that the data in Table 15 are largely exploratory. It would be valuable to further explore the differential performance of students with disabilities on the measures.

Table 15. Mean correlations between single and aggregated probe scores and students' scores on other indicators of algebra proficiency.

		Algebra Grade	Teacher Rating	IAAT Total	ITED Con/App	ITED Comp	ITED Reading
Basic Skills	5 min.						
Single Probes		.74	.68	.71	.10	.60	.44
Aggregated		.77	.71	.67	.05	.56	.31
Algebra Foundations	5 min.						
Single Probes		.42	.36	.00	-.35	-.48	-.41
Aggregated		.53	.48	.11	-.32	-.21	-.29
Translations	Correct	.51	.28	.76	.70	.85	.32
Single Probes	C – I	.12	.59	.31	.79	.86	.26
	C – 1/3	.40	.68	.62	.77	.85	.29
Translations	Correct	.51	.28	.76	.70	.84	.32
Aggregated	C – I	.34	.59	.63	.72	.83	.73
	C – 1/3	.43	.57	.77	.71	.76	.71
Content Analysis	Correct	.35	.33	-.03	-.25	-.15	-.27
Multiple Choice	C – I	.36	.37	.03	-.23	-.09	-.16
Single Probes	C – 1/3	.35	.34	.00	.24	-.13	-.23
Content Analysis	Correct	.34	.33	-.03	-.25	-.15	-.27
Multiple Choice	C – I	.48	.48	.14	-.26	-.10	-.11
Aggregated	C – 1/3	.47	.44	.11	-.26	-.10	-.15

Summary and Considerations for Future Research

The purpose of this study was to examine the technical adequacy of four potential measures of algebra proficiency. Seventy-eight students in grades nine to twelve from two different Iowa school districts participated in the study; 6 of these students were receiving special education services. The data were gathered from March through May 2005, 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 for each of four types of algebra probes: *Basic Skills*, *Algebra Concepts*, *Translations*, and *Content Analysis-Multiple Choice*. The testing sessions were spaced one week apart and were preceded one to 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 four 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 about half 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 varied a great deal across the four types of algebra probes. Increases in reliability (over results obtained in previous studies, Foegen & Olson, 2005; Foegen, Olson, & Perkmen 2005) were noted for the *Basic Skills* and the *Content Analysis-Multiple Choice* measures. The reliability of single *Basic Skills* probes was at or above the standard benchmark of .80 in most instances. For the *Content Analysis-Multiple Choice* measure, reliability estimates for single probes increased from the .4 to .5 range in earlier studies to the .6 to .8 range in the present study. Small increases were obtained for the *Algebra Foundations* measure, with mixed results (some increases and some decreases) for the *Translations* probes.

When scores from two probes were aggregated, all reliability estimates increased to the .7 to .8 range. In comparison to previous results, the aggregated reliability estimates were higher for all probes except the *Translations* measure. For the present study, we introduced the use of a sample page prior to the first administration of each type of probe. We also made revisions to the design templates for some of the probes. Our data suggest that these actions had a positive effect on the reliability of our measures. With the exception of the *Basic Skills* measure, the remaining probes continue to fall short of standard expectations for reliability in assessments that will be used to make educational decisions (including monitoring student progress). Further research should examine the effect of on-going administration of the measures to determine whether student performance becomes more reliable as a function of familiarity with the task. In the present study, students completed two forms of each task on each of two different occasions. It may require more exposures to these algebra tasks in order to obtain sufficiently stable estimates of student performance levels.

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). With regard to the grade-based measures and teacher ratings, criterion validity coefficients for single probes were in the low to moderate range. Overall GPA was not strongly associated with any of the algebra measures. The *Basic Skills* measure had the strongest relations with these criterion variables, followed by the *Content Analysis-Multiple Choice* measure. The *Algebra Foundations* measure had somewhat lower coefficients and the *Translations* measure was unrelated to teacher ratings and had low coefficient with algebra grades. Aggregating scores produced moderate increases in the strength of the coefficients. The general pattern of results remained the same.

The district's achievement test (ITED) was not related to students' performance on the algebra measures. This result was not surprising, as the ITED has minimal algebra content. In general, student performance was more associated with students' scores on the Computation subtest than on the Concepts/Problems subtest. No relations with the reading subtest were identified for any of the algebra measures.

When we examined the correlations between students' scores on the IAAT, the *Basic Skills* and *Algebra Foundations* measures demonstrated the strongest relations, followed by the *Content Analysis-Multiple Choice* measure. The *Translations* measure produced the lowest coefficients. Aggregating score produced only slight improvements, with the overall pattern of results remaining the same.

We also conducted exploratory analyses for the subset of students with disabilities in the larger sample. Although this group was small (only 6 students), we were curious to see whether the same pattern of results would hold for these students as for the larger sample. Our findings, which must be considered with great caution, were surprising. The probes that demonstrated the highest levels of criterion validity for students with disabilities were the *Basic Skills* measure and the *Translations* measure. Although these findings clearly must be replicated with a larger group of students, we hypothesize that the *Translations* measure may allow students with disabilities who struggle with the procedural aspects of traditional algebra instruction to have a means to demonstrate their understanding of conceptual features of algebraic relationships.

<u>Considerations for Future Research</u>

Several issues arose during this study that should be addressed in future research. First, we have concerns about the viability of the *Translations* measure for future work in Project AAIMS. Although we hold out hope that the measure may prove to be useful at some point in the future, we have not found it to have sufficient technical adequacy for the classrooms participating in our project. Given that all three districts are using a more traditional basal series, the conceptual nature of the *Translations* probe generates great difficulty for a majority of the students. Our exploratory analyses of the students with disabilities suggest that this measure may be effective for particular groups of students. Future research should include smaller-scale studies that examine the performance of students with disabilities and other low achieving students on this measure. It would also be appropriate to identify settings in which more conceptual approaches to instruction are being used and examine the technical adequacy of the measure in those settings.

Second, the revisions to our measures and administration procedures that were implemented in this study appear to have had positive results. In general, the reliability and the criterion validity of the measures explored here were higher than those obtained in previous studies in similar settings. Future research should continue to explore minor revisions to the measures that might further enhance their technical adequacy. As one example, our error analyses of students' work on the *Content Analysis-Multiple Choice* measure indicated that very few students even attempted the final four problems on this probe, which corresponded to the later chapters of the text. Future revisions might include adapting the design template for this measure to focus more attention on earlier chapters. In addition, we often found that as students completed this measure, they stopped as soon as they encountered a difficult problem. This was likely based on their reasoning that if the problem were ordered parallel to the chapters of the text (which they were), that the problems from subsequent chapters would be more difficult, and therefore there was no point in even attempting additional problems. While this reasoning is logical, there were other students who did consider all the problems and found some from later chapters that they were able to solve. Future revisions to the *Content Analysis-Multiple Choice* measure might also include randomly ordering the problems on the probe and encouraging students to consider each of the problems to identify those they know how to solve.

Finally, we occasionally encountered a few 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:

- Revising the design template for the *Content Analysis-Multiple Choice* measure to shift the balance of sampling from the text so that more problems are drawn from earlier chapters
- Altering the construction procedures so that problems on the *Content Analysis-Multiple Choice* measure are in random order, rather than in chapter order
- Exploring (with teachers) means of engaging students in the probe completion process so that more confidence can be placed that their scores represent their best work
- Conducting smaller scale studies on the *Translations* measure that explore its viability for use with students who are struggling to learn algebra, as well with typical students who are receiving instruction in a setting that emphasizes conceptual understanding over procedural proficiency

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.

Foegen, A., Olson, J., & Perkmen, S. (2005). *Reliability and criterion validity of five algebra measures in Districts B and C* (Tech. Rep. No. 7). 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

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

Solve: $9 + a = 15$ $a =$
Evaluate: $12 + (-8) + 3$
Simplify: $2x + 4 + 3x + 5$
Solve: $12 - e = 4$ $e =$
Simplify: $4(3 + s) - 7$
Simplify: $b + b + 2b$
Solve: $6 \bullet 9 = d$ $d =$
Simplify: $7 - 3(f - 2)$
Evaluate: $(-5) + (-4) - 1$
Solve: $63 \div c = 9$ $c =$
Simplify: $17 + 6d + 2d - 9$
Simplify: $8m - 3(m - 2)$
Solve: $y + 4 = 11$ $y =$
Evaluate: $4 - (-2) + 8$
Simplify: $2k + 8 - 5(k + 7)$

Solve: $12 - 6 = g$ $g =$
Simplify: $4d + 7d - 9 + 2$
Simplify: $5(b - 3) - b$
Solve: $4r = 28$ $r =$
Evaluate: $8 - (-6) - 4$
Simplify: $w - w(4 + 5) - 6$
Solve: $\frac{z}{5} = 5$ $z =$
Simplify: $4 - 7b + 5(b - 1)$
Simplify: $s + 2s - 4s$
Solve: $6 + 7 = v$ $v =$
Simplify: $5(q + 3) - 9$
Evaluate: $9 + (-3) - 8$
Solve: $r - 4 = 4$ $r =$
Simplify: $y^2 + y - 4y + 3y^2$
Simplify: $3(c + 2) - 2c$

Solve: $3 \bullet 8 = m$ $m =$
Evaluate: $-9 + 3 + 8$
Simplify: $x + 2(x - 5) - 3$
Solve: $\frac{36}{6} = s$ $s =$
Simplify: $5(3 + f) - 2f + 6$
Simplify: $5 - 2b + 4(b + 3)$
Solve: $r + 7 = 15$ $r =$
Simplify: $4(y + 1) - 8y$
Evaluate: $14 - 7 + (-3)$
Solve: $12 - d = 4$ $d =$
Simplify: $-3w^2 + 5w^2 - 5 + 12$
Simplify: $9 - 4(v - 2)$
Solve: $5q = 30$ $q =$
Simplify: $16 + 2(t - 4) - 3t$
Evaluate: $-2 + (-5) + (-8)$

Solve: $28 \div 4 = d$ $d =$
Simplify: $7b - 4 - 3 - 2b$
Simplify: $2e - 3(e - 4)$
Solve: $4 + 7 = x$ $x =$
Evaluate: $-5 + 6 - 6$
Simplify: $4 + 10(1 - r)$
Solve: $18 - 9 = k$ $k =$
Simplify: $6r - 5 - 2r + 6$
Evaluate: $-1 + (-4) + (-7)$
Solve: $5 \bullet 7 = j$ $j =$
Simplify: $3(u + 3) - 2u + 5$
Simplify: $2c - 3c - c$
Solve: $\frac{h}{6} = 8$ $h =$
Simplify: $a(7 - 2) - 2a + 4$
Simplify: $3z - 8z + 2 + 9$

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

Solve:
$y \bullet 8 = 48$
$y =$

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

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

Solve:
$e - 5 = 4$
$e =$

Simplify:
$3a + 5a - a$

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

Solve:
$42 \div s = 6$
$s =$

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

Simplify:
$r + 5(r + 1) + 9$

Solve:
$7 + b = 12$
$b =$

Simplify:
$8 - 5(y + 1)$

Simplify:
$4(a + 2) - 3a$

Solve:
$9q = 81$
$q =$

Simplify:
$7 - 2d + 3(d + 6)$

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

Solve:
$14 - e = 9$
$e =$

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

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

Solve:
$7 \bullet 3 = k$
$k =$

Evaluate:
$1 - 9 + (-2)$

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

Solve:
$s - 5 = 6$
$s =$

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

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

Solve:
$z + 7 = 13$
$z =$

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

Simplify:
$-2(n - 6) + 5n$

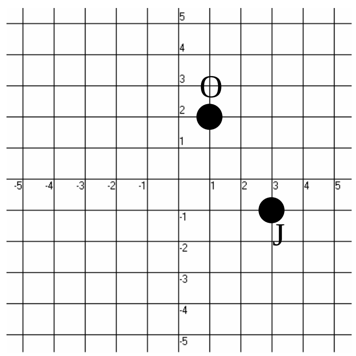
Solve:
$\frac{63}{9} = s$
$s =$

Simplify:
$5 - 2b + b(4 + 3)$

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

Find the ordered pair for each point:

J(,) O(,)



Fill in the empty box:

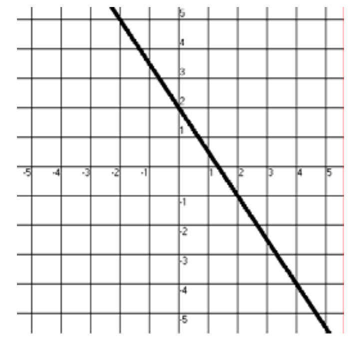
s	$3s$
6	18
7	21
8	
9	27

Fill in the empty box:

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

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 _____

Evaluate:
$9 \bullet 4 - 6$

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

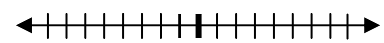
Solve:
$n + 3 = 8$
$n =$ ________

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

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

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

Graph the expression $m > 6$



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

Evaluate:
$8 \div 2 + 4 \bullet 3$

Evaluate:
2^3

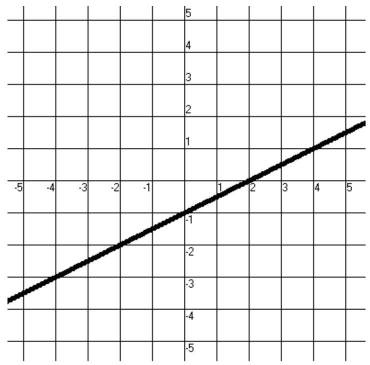
Write the expression for this phrase:
9 times a number

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

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

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:

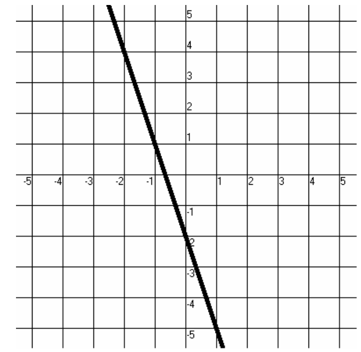
n	
6	4
9	6
12	8
15	10

Fill in the empty box:

t	t - 7
-2	-9
2	-5
6	
10	3

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:
$x \div 4$

Evaluate:
$(-16) \div (-4)$

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

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

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

Graph the expression $p \leq -3$

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

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

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

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

Evaluate:
$10 - 3 \bullet 5 + 2$

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

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

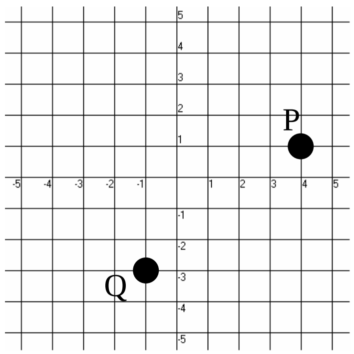
Evaluate:
$(-3)(9 - 7)$

Evaluate:
$\sqrt{36}$

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

Find the ordered pair for each point:

P(,) Q(,)



Fill in the empty box:

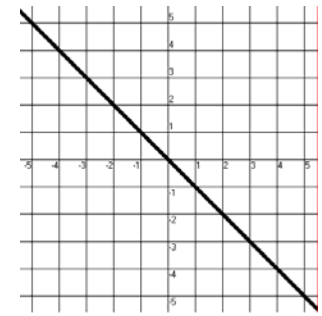
t	$4t + 1$
-1	-3
1	5
3	
5	21

Fill in the empty box:

y	
5	1
10	2
15	3
50	10

Fill in the empty box:

n	$n+3$
-2	1
-4	-1
-6	
-8	-5



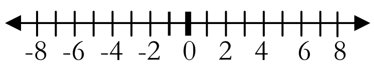
What is the slope?

What is the *y* intercept?

Write the expression for this phrase:
a number divided by 4

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

Graph the expression $t < 2$



Evaluate:
$7 – 15 \div 3$

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

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

Evaluate:
$(– 6) \bullet (– 8)$

If $y < 5$, two possible values for *y* are _________ and _________

Evaluate:
$(– 5)(8 – 6)$

Evaluate:
$\sqrt{49}$

Simplify:
$8t + (3t – t)$

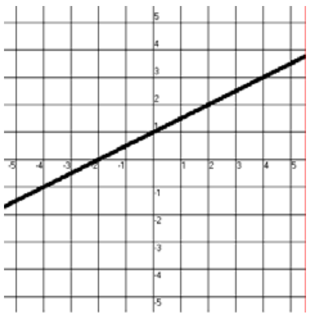
Solve:
$9x = 45$
$x =$ _________

Evaluate:
$4 \bullet 5 – 2 + 6$

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

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

Evaluate:
$2 + 2 \bullet 4 – 4$



What is the slope?

What is the y intercept?

Fill in the empty box:

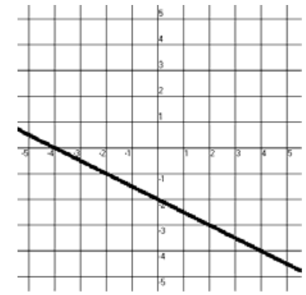
x	
5	20
10	40
15	60
20	80

Fill in the empty box:

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

Fill in the empty box:

y	
-5	2
1	8
5	12
10	17



What is the slope?

What is the y intercept?

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

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

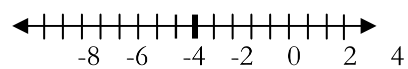
Simplify:
$12n – 7 + 3n + 4$

Write a word phrase for this expression:
$5t + 2$

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

Evaluate:
3^3

Graph the expression $p \geq – 5$



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

Solve $18 – n = 12$
$n =$ _________

Evaluate
$– 4(8 + 2)$

Simplify:
$5b – (b + 6b)$

Evaluate:
$8 \div 2 + 6 \bullet 2$

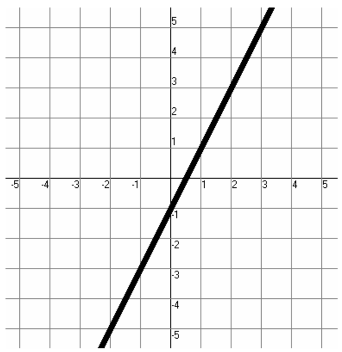
Write a word phrase for this expression:
$\frac{18}{b}$

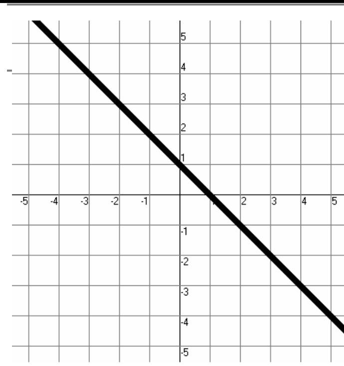
Evaluate $5x – 4$ when
$x = 4$ _________
$x = 8$ _________

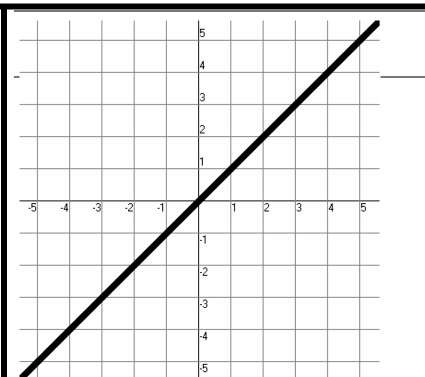
Write the expression for this phrase: *3 times a number*

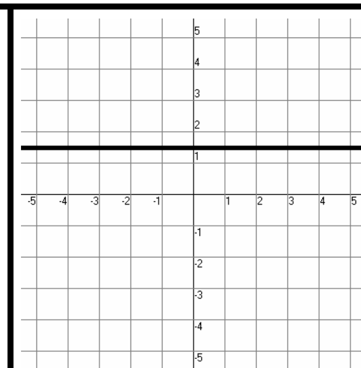
Simplify:
$5(m + 2) – 3m$

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









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

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

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

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

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

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

A

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

B

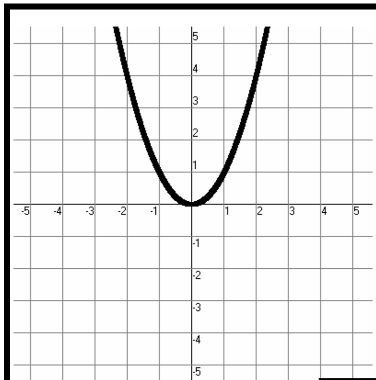
x	y
2	4
1	2
0	1
-1	$\frac{1}{2}$
-2	$\frac{1}{4}$

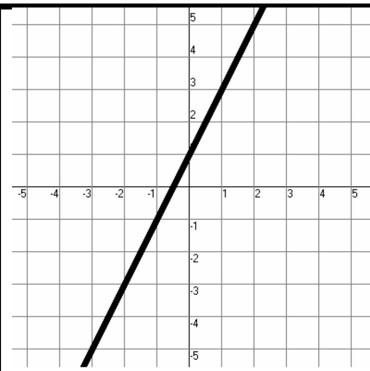
C

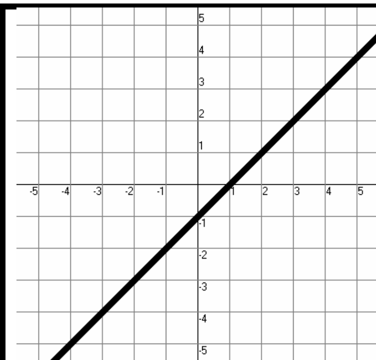
x	y
2	5
1	3
0	1
-1	-1
-2	-3

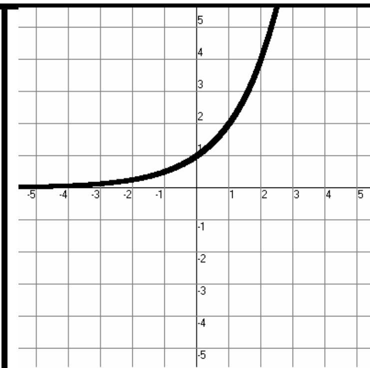
D

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







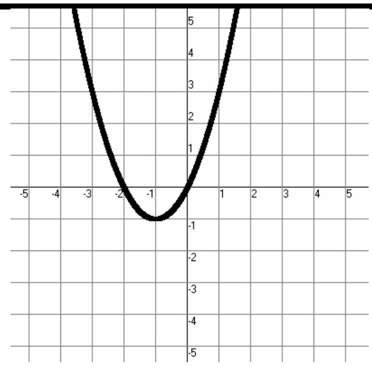


$y = 2x + 1$ ____	$y = 2^x$ ____	$y = x - 1$ ____	$y = x^2$ ____	$3y = 6x + 3$ ____

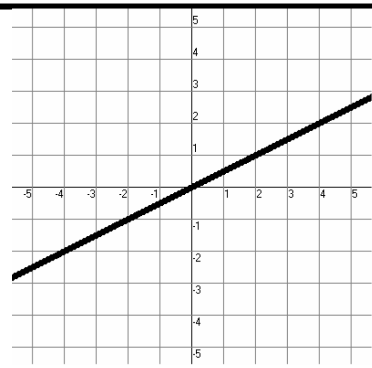
Mr. Jones is going to give a true/false test. He made this data table to show the number of possible answer combinations his students can give on the test.	____
Sue made this data table to figure out how many inches of wire she needs for a bracelet. Each bracelet uses two strands and she needs to add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

ALGEBRA PROBE D-11

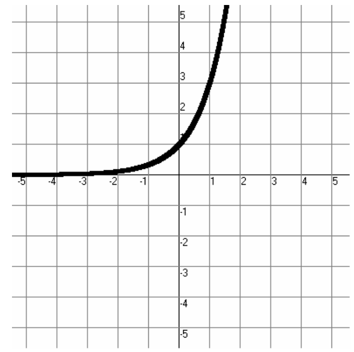
A



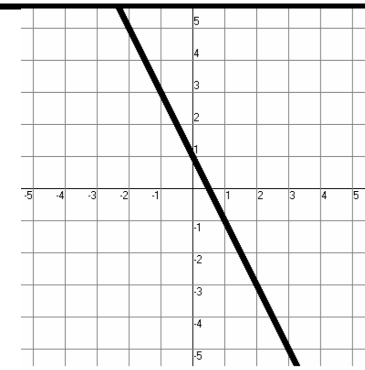
B



C



D



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

____

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

____

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

____

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

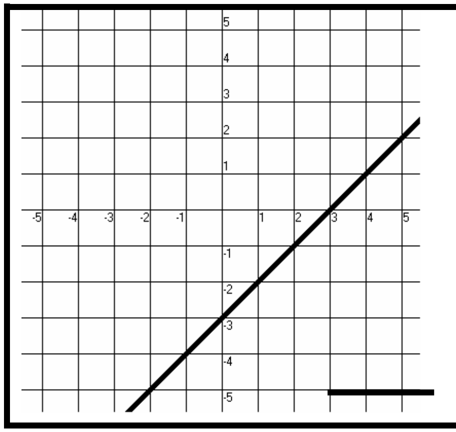
____

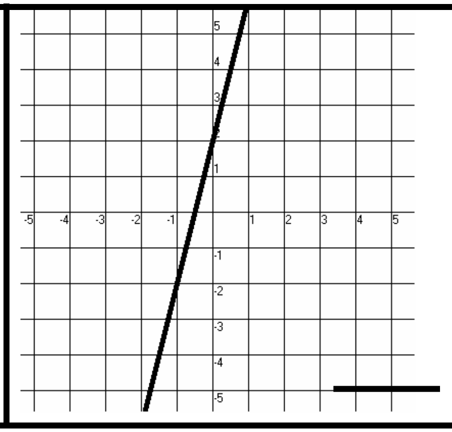
$y = -2x + 1$ ____	$y = x^2 + 2x$ ____	$y = 3^x$ ____	$y = x(x + 2)$ ____	$y = \frac{1}{2}x$ ____

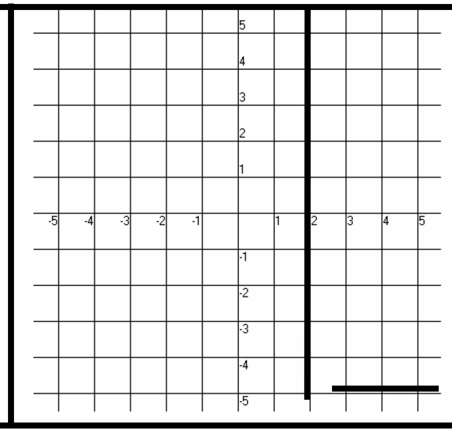
Matt built a maze for his gerbil. Each time the gerbil comes to an intersection, it can go three possible ways. Matt made this graph to show the total possible number of routes for the gerbil through the maze.	____
LaShaya's mom makes her save half of what she earns in the summer for college. She made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

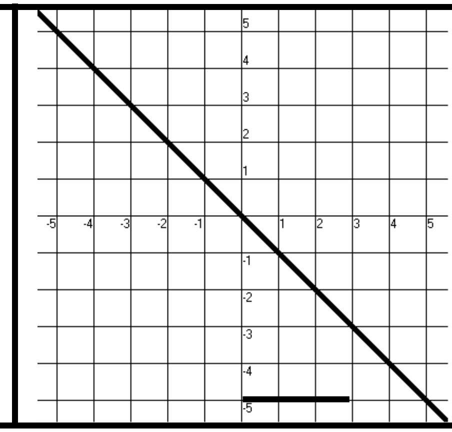
PROBE D-12

A	B	C	D
$y = 4x + 2$	$y = x - 3$	$y = -x$	$x = 2$









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 wrote this equation to show how many quarters he'll have after the trade.	____
Leah is three years younger than her sister. She wrote this equation to show the relationship between their ages.	____
Every time he gets home after curfew, he loses a chance to use the car. Joel wrote this equation to show the relationship between breaking curfew and his chances to use the car.	____
Sam is planning a basketball tournament. He wrote this equation 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 wrote this equation to show how her total quiz points would be related to the score she gets on each quiz.	____

A

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

B

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

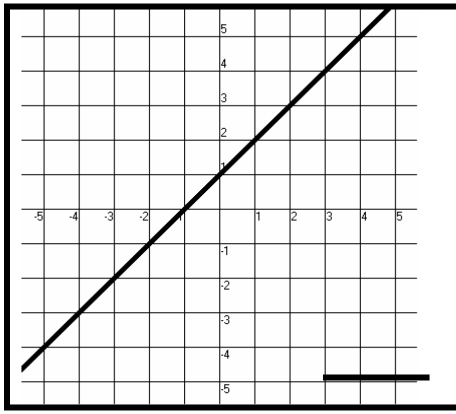
C

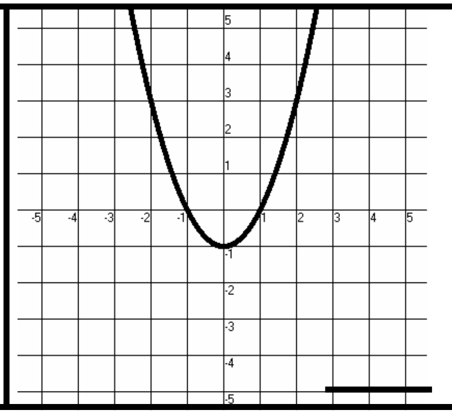
x	y
2	4
1	8
0	16
-1	32
-2	64

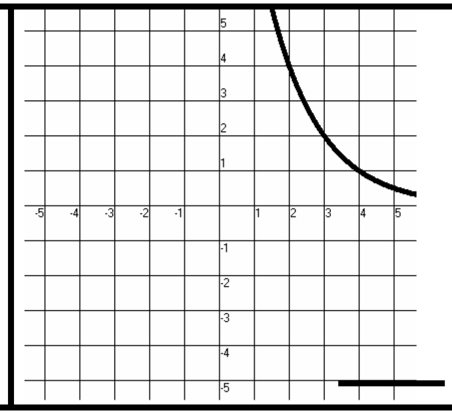
D

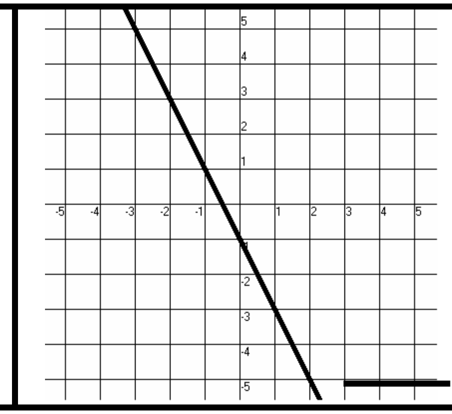
x	y
2	3
1	2
0	1
-1	0
-2	-1

$y = 16(.5)^x$ ____	$y = -2x - 1$ ____	$4y = -8x - 4$ ____	$y = x + 1$ ____	$y = x^2 - 1$ ____



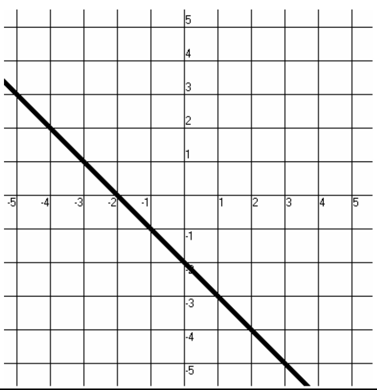




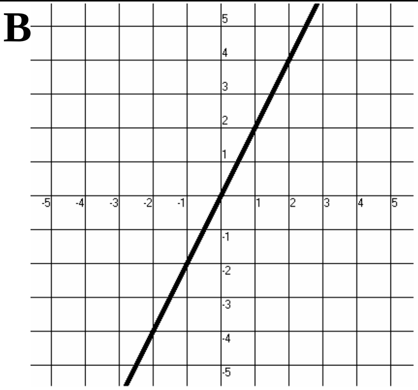


Pat is organizing the brackets for the doubles tennis tournament. Sixteen teams have entered. Pat made this data table to show how many teams will be left after each of the rounds.	____
LeRoy needs to buy tile for a square room. The tiles come in 1-foot squares. There is a post in the middle of the room that is the same size as one tile. LeRoy made this data table 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 made this data table 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 made this data table 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 made this data table to find the height of any person who stands on the scale.	____

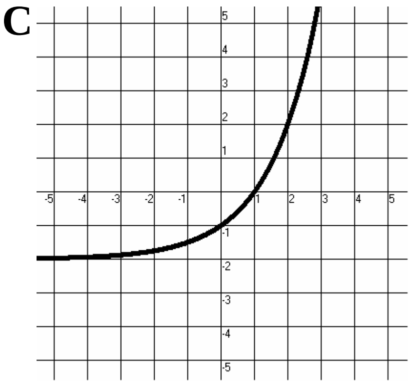
A



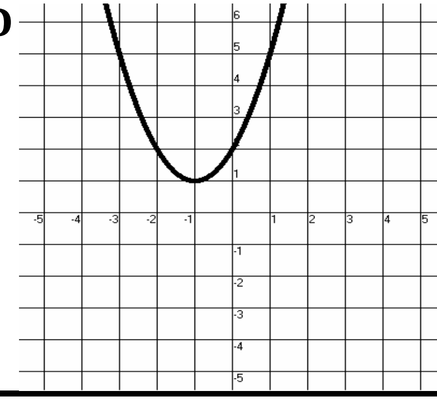
B



C



D



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

____

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

____

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

____

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 graph 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 graph 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 graph 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 graph 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 graph shows the relationship between Jean's old and new hourly pay rates.	____

Evaluate $a^2 - b \div 2$ when $a = 4$ and $b = 6$

a) 1
b) 5
c) 10
d) 13

Simplify:
$3(m + 2) + 2(m - 1)$

a) $5m + 4$
b) $5m + 1$
c) $6m + 8$
d) $6m - 8$

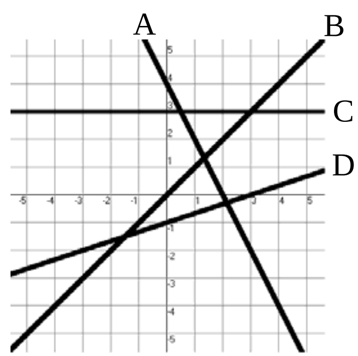
Simplify:
$6(2b - 3) - 3(2 - b)$

a) $15b - 24$
b) $9b - 9$
c) $9b + 12$
d) $15b + 12$

Solve:
$6c + 4 = -3c - 14$

a) $-\frac{10}{3}$
b) -2
c) 2
d) 6

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



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

Find the slope of a line through (1, –1) (5, 2)

a) $\frac{1}{5}$
b) $\frac{3}{4}$
c) -6
d) $-\frac{4}{3}$

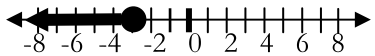
Write the equation in slope-intercept form if $m = \frac{1}{2}$ and $b = 3$

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

Write the equation of a line through (5, 3) (4, 9). Use point-slope form.

a) $y + 1 = 2(x - 4)$
b) $y + 4 = -6(x - 1)$
c) $y - 3 = -6(x - 5)$
d) $y = -6x + 30$

This graph shows the solution for which equation?



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

Which of the following is a logical first step to solve this linear system using substitution?

$3x + 2y = 4$
$4x + y = 7$

a) $x = -2y + 4$
b) $y = -4x + 7$
c) $4 = 3x + 2y$
d) $4x = -y + 7$

Solve the linear system:

$x - y = 4$
$x + 2y = 19$

a) $(-1, -5)$
b) $(5, 8)$
c) $(-2, 19)$
d) $(9, 5)$

Simplify, with no negative exponents:

$$\left(\frac{6x^2y^{-1}}{2xy}\right)^2$$

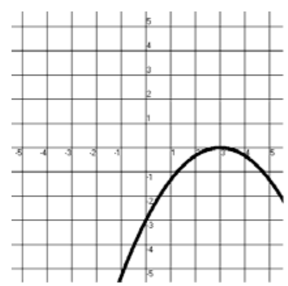
a) $9x^2$ b) $3x^2y^3$
c) $\frac{3x}{y^2}$ d) $\frac{9x^2}{y^4}$

Simplify:

$$\sqrt{32}$$

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

Which function matches this graph?



a) $y = \frac{1}{3}x^2 - 7x - 2$
b) $y = x^2 + 2x + 3$
c) $y = -\frac{1}{3}x^2 + 2x - 3$
d) $y = -x^2 - 3$

Factor this trinomial:

$2x^2 + 5x - 3$

a) $(x - 2)(x - 1)$
b) $(2x - 1)(x + 3)$
c) $(2x + 1)(x - 3)$
d) $(x - 1)(x + 3)$

Simplify:

$$\frac{12}{2x+4} + \frac{3x}{x+2}$$

a) $\frac{3x+12}{3x+6}$ b) $\frac{x+4}{x+2}$
c) 3 d) $9x$

PROBE E-22

Evaluate $2x + 3 - y^3$ when $x = 5$ and $y = 2$

a) –3
b) 5
c) 7
d) 21

Simplify:
$4(s – 2) + 5(s + 3)$

a) $s + 1$
b) $9s – 5$
c) $9s + 1$
d) $9s + 7$

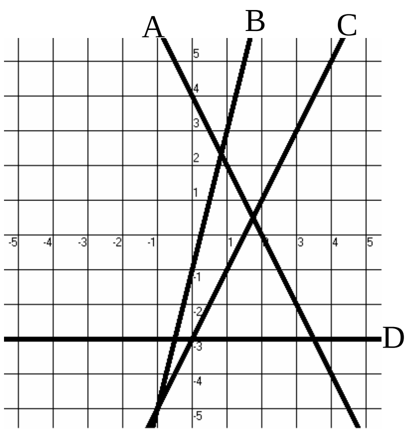
Simplify:
$2(z + 5) – 3(z – 2)$

a) $z + 16$
b) $z + 3$
c) $–z + 4$
d) $5z + 6$

Solve:
$9r – 2 = 24 – 4r$

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

Which line on the graph is
$y – 4x = –1$?



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

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

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

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

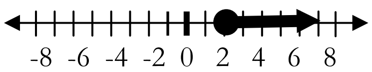
a) $y = 3x + 2$
b) $3y = 3x + b$
c) $y = 2x – 3$
d) $y = 3x + 4$

Write the equation of a line through (–2, –8) (2, 4). Use point-slope form.

a) $y – 4 = 3(x – 2)$
b) $y + 8 = 5(x + 2)$
c) $y – 3 = 3(x – 5)$
d) $y = 3x - 2$

PROBE E-22

This graph shows the solution for which equation?



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

Which of the following is a logical first step to solve this linear system using substitution?

$2x = 6y - 4$
$3x - 5y = -10$

a) $x = \dfrac{5y - 10}{3}$
b) $5y = 3x + 10$
c) $x = 3y - 2$
d) $4 = 6y - 2x$

Solve the linear system:

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

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

Simplify, with no negative exponents:

$$\left(\frac{8s^{-2}t^{2}}{4t}\right)^{3}$$

a) $\dfrac{2t}{s^2}$ b) $\dfrac{8t^3}{s^6}$

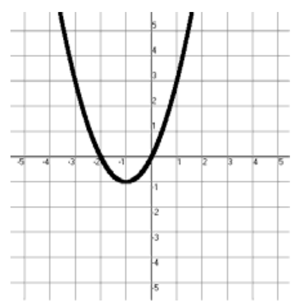
c) $\dfrac{2s^2}{t}$ d) $2s^2t$

Simplify:

$$\sqrt{75}$$

a) $3\sqrt{5}$
b) $\sqrt{25} \bullet \sqrt{3}$
c) $25\sqrt{3}$
d) $5\sqrt{3}$

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:

$$\frac{3x}{x^2 - 4} - \frac{x + 4}{x^2 - 4}$$

a) $\dfrac{3x^2 + 4}{x^2 - 4}$ b) $\dfrac{4x + 4}{x^2 - 4}$

c) $\dfrac{2}{x + 2}$ d) $\dfrac{2x + 4}{x^2 - 4}$

Algebra Probe Data Collection Procedures
District B Senior High
District C Senior High
April and May 2005

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. We will be back in your class to do some more of the probes next week. Thank you for your help with our research project!***

<u>Directions for the Algebra Basic Skills Probes</u>: **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.***

<u>Directions for Translations Probes: **D1, D2**</u>

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.***

<u>Directions for Content Analysis-Multiple Choice Probes: **E1, E2**</u>

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 ___________________

Project AAIMS:
Algebra Assessment and Instruction:
Meeting Standards
District B Senior High

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