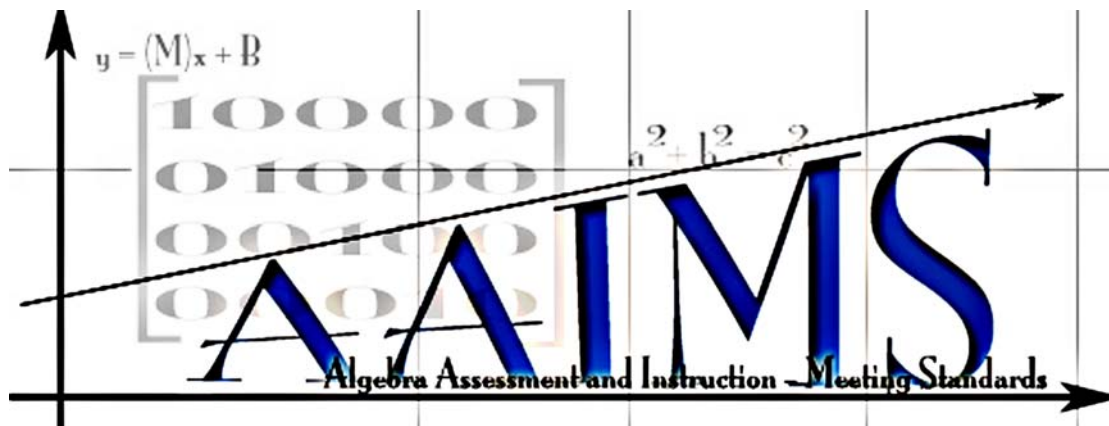


**PROJECT AAIMS: ALGEBRA ASSESSMENT AND  
INSTRUCTION – MEETING STANDARDS**



A Replication Study of the Reliability, Criterion Validity and Sensitivity  
to Growth of Two Algebra Progress Monitoring Measures

Technical Report #13

Serkan Perkmen, M.Ed.

Anne Foegen, Ph.D.

Jeannette Olson, M.S.

Iowa State University

September 2006

PROJECT AAIMS: ALGEBRA ASSESSMENT AND  
INSTRUCTION – MEETING STANDARDS

A Replication Study of the Reliability, Criterion Validity of Two Algebra  
Progress Monitoring Measures

Technical Report 13

---

**Abstract**

---

This study served as a replication of previous work examining the reliability, validity, and sensitivity to growth of the Basic Skills and the Content-Analysis-Multiple Choice probes in two Iowa schools districts. One hundred five students in grades nine to twelve participated in the study. Data were gathered from February 2006 to April 2006. Over three months of data collection, students completed two Basic Skills probes and two Content Analysis-Multiple Choice probes each month. We examined the alternate form reliability and test-retest reliability for both types of probes. We found that both types of probes possessed adequate levels of reliability, with the Basic Skills probes demonstrating a higher level of reliability than the Content Analysis-Multiple Choice probes. To assess the validity of the probes, we gathered data from a variety of indicators of students' proficiency in algebra including course grades, teachers' evaluations of student proficiency and growth, and performance on standardized assessment instruments including the Iowa Test of Education Development (ITED) and the Iowa Algebra Aptitude Test (IAAT). We examined both concurrent and predictive validity. Concurrent validity was supported by finding moderate correlations between both types of probes and criterion measures including teachers' evaluation of their students' proficiency and IAAT scores. Predictive validity of both types of probes was supported by finding low or moderate correlations between the relationship of the earliest probe scores to other indicators administered at the end of the course including teacher ratings of growth, students' end-term algebra grades, IAAT scores taken at the end of the semester, and ITED scores. We also examined the extent to which both types of probes reflect student growth and explored the relationship between student growth on the algebra measures and other indicators of growth. We found that students who did not drop the course grew .42 and 1.52 points each week on Basic Skills and Content Analysis- Multiple Choice probes, respectively. This result suggests that Content Analysis-Multiple Choice probes may be more sensitive to reflecting student growth. We also found that there was a small but significant correlation between teacher ratings of growth and Content Analysis-Multiple Choice probes. No significant correlation existed between Basic Skills probes and other indicators of growth.

---

## Full Report

---

### Introduction

Previous work in Project AAIMS has established the reliability and criterion validity of three measures for monitoring student progress in algebra. In Technical Report 10, we reported the technical features of the measures when used for static measurement of student performance. The three measures (Basic Skills, Algebra Foundations, Content Analysis-Multiple Choice) have acceptable levels of reliability and moderate levels of criterion validity. Technical Report 12 described the results of our initial studies investigating the degree to which the measures may be valuable for progress monitoring purposes. In this study, we learned that the both the Basic Skills and the Content Analysis-Multiple Choice probes were able to detect changes in student performance over time. The Content Analysis-Multiple Choice measure proved to be slightly more sensitive (i.e., produced higher weekly slope values) than the Basic Skills measure. Both probes resulted in student gains that met or exceeded a .5 units/week threshold. The present study serves as a replication of the earlier study.

### Method

The study described in this report was conducted from February to April 2006 in Districts B and C. District B is located in a community of 26,000 people, where the high school currently enrolls 1,349 students. The majority of students in District B are white (82%), and nearly half are eligible for free and reduced lunch (47%). Eighteen 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 472 students in grades 9 through 12. In District C, 39% of the students are of diverse backgrounds in terms of race, culture and ethnicity. Approximately 45% of the district 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 from February to April 2006. During the first data collection session, students completed the algebra criterion measure. All data collection activities involving students were completed during regular class time. Teachers administered all algebra probes.

### Participants

Sixty students in District B and 45 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	60	41	15	3	1
Gender					
Male	28	17	9	2	0
Female	32	24	6	1	1
Ethnicity					
White	55	38	13	3	1
Black	4	2	2	0	0
Hispanic	1	1	0	0	0
Lunch					
Free/Red	20	12	8	0	0
Disability					
IEP	2	0	2	0	0

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

	Total	Grade 9	Grade 10	Grade 11	Grade 12
N	45	30	8	6	1
Gender					
Male	18	14	3	1	0
Female	27	16	5	5	1
Ethnicity					
White	31	21	6	3	1
Black	0	0	0	0	0
Hispanic	7	4	1	2	0
Native Am.	7	5	1	1	0
Lunch					
Free/Red	20	12	4	3	1
Disability					
IEP	5	1	1	3	0

As the data in Tables 1 and 2 indicate, many of the participants (an average of 82%) were white and an average of 68% were in ninth grade, the traditional grade in which students in these districts complete algebra. Thirty-three and 44 percent participated in federal free or reduced lunch programs in Districts B and C, respectively and 3% and 13% of the participating students in Districts B and C, respectively, were students with disabilities who were receiving special education services. In District B, 20 students were enrolled in Algebra 1A, 40 in Algebra 1. Algebra 1A and Algebra 1B is an option available in District B in which students complete half the content of a traditional Algebra 1 course in a single course. A similar arrangement was offered in District C in which a small number of students were enrolled in an algebra course that was only half the duration of a typical block class (45 minutes) and lasted for the entire school

year. For simplicity, we use the same Algebra 1A/1B terms from District B to refer to this option in both districts. In District C, 7 students were enrolled in Algebra 1B and 38 students were in Algebra 1. Due to the relatively small numbers of students in each district participating in the study, data from students in the two schools were combined for statistical analyses purposes.

Additional Information on Students with Disabilities. Because the applicability of the algebra probes to students with disabilities is an important part of Project AAIMS, additional information about the 7 students with disabilities in District B and C participating in the project is provided in Table 3.

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

Characteristic	Quantification
Disability category	2 Entitled Individual (EI) 4 Learning Disability 1 Deaf/Hard of Hearing, Severe/Profound
% time in general education	Range = 78 –100%; Mean = 87.5%
# of students with math goals	2
# of students receiving math instruction in general education classes	7

In algebra, students with disabilities earned mean grades of 1 [D] (range 0.00 [F] to 2.00 [C]). In Districts B and C, the Iowa Tests of Educational Development are used as a district-wide assessment. On average, students with disabilities obtained national percentile rank scores of 28 and 48 in Concepts/Problem Solving and Computation, respectively.

Algebra Progress Monitoring Measures. Two 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 two types of algebra measures

*Probe A: Basic Skills Measure*

Probe A is 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 hypothesize 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 address solving simple equations, applying the distributive property, working with integers, combining like terms and applying proportional reasoning. The *Basic Skills* probe includes many skills one would assume that students proficient in algebra would be able to complete with reasonable levels of automaticity. Students have five minutes to work on this probe; six parallel forms were used in the study. Each *Basic Skills* probe consists of 60 items; each item is scored as one point if it is answered correctly.

### *Probe E: Content Analysis-Multiple Choice Measure*

The *Content Analysis-Multiple Choice* measure is a variation of an earlier measure that used a constructed response format (see Technical Report 7 for additional details on this measure). Our rationale for including a multiple choice option is to see if this format improves scoring efficiency (and potentially interscorer agreement). Another goal is to familiarize students with the multiple-choice format used on district-administered assessments. Like the constructed-response version, this probe consists of 16 items that correspond to different chapters in the textbook that is used in all three districts. Unlike the constructed-response version, this probe only includes problems from chapters 1 to 8, rather than from the entire textbook. There are six parallel forms for this probe. Students have seven minutes to work on the Content Analysis-Multiple Choice probes.

Scoring for the *Content Analysis-Multiple Choice* probes is done by comparing student responses to a rubric-based key created by the research staff. Each of the 16 problems is worth up to three points. Students earn full credit (three points) by circling the correct answer from among the four alternatives. If students circle an incorrect response and do not show any work, their answer is considered a ‘guess;’ the total number of guesses is recorded for each probe. In cases where students show work, the scorer compares the student’s work to the rubric-based key, and determines whether the student has earned 0, 1, or 2 points of partial credit. The number of points earned across all 16 problems and the number of guesses are recorded and entered in the data files. A final score is computed by subtracting the number of guesses from the total number of points earned on the probe.

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

The classroom-based measures included grade-based measures and teacher ratings. Each student’s *algebra grade*, the grade s/he earned in algebra during the spring semester of the 2005-2006 school year, was recorded using a four-point scale (i.e., A = 4.0, B = 3.0). In District B, students earned a course grade at the end of the spring semester because they had completed a full course in that semester due to the block scheduling format. In District C, students earned grades at four points during the school year (two of which fell during the midpoint and end of the present study). For District C students, we averaged the grades they earned during the duration of the study and used this average as their course grade.

We also included the teachers’ evaluations of student proficiency in algebra by asking each teacher to complete a *teacher rating of proficiency* for all the students to whom she/he taught algebra. These ratings were completed at the beginning of the course. Student names were alphabetized across classes to minimize any biases that might be related to 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. This enabled us to see if there was a relationship between growth that students showed on both types of probes and the teachers’ evaluations of student growth.

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. We recorded the Concepts/Problems subtest score (which was identical to the Math Total score) and the Computation subtest score.

Because the district-administered measure did not provide a direct assessment of algebra, we also administered the *Iowa Algebra Aptitude Test (IAAT)*. This norm-referenced instrument is typically used to evaluate the potential of 7<sup>th</sup> grade students for successful study of algebra in 8<sup>th</sup> grade. We recorded and use national percentile rank scores for the total scale for our validity analyses to parallel the data available for the ITED. 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.

Growth Measures. One of the major goals of the AAIMS project is to determine the extent to which the two types of measures reflect student growth over time. We were also interested in exploring whether the growth that students showed in on the probes is associated with other indicators of growth. To accomplish these goals, we gathered data using three types of measures reflecting students' growth: *probe slope*, *teacher rating of growth*, and *IAAT gain score*.

The first type of growth measure, which we called *probe slope* reflects the growth that students' showed on both types of probes over the semester. We used ordinary least square regression to calculate each student's slope on each measure. The obtained slope values were calculated to reflect the amount of weekly progress a student demonstrated on a probe type. The second type of measure was the *teacher rating of growth*. At the end of the semester, we asked teachers to rate all the students in their algebra classes. Student names were alphabetized across class periods to minimize any biases that might be related to particular class sections. Teachers used a 5-point Likert scale to rate each student's growth in algebra in comparison to same-grade peers. A rating of 1 indicated minimal or no growth, while a rating of 5 represented unusually high growth in comparison to peers. The third type of measure was the *IAAT gain score*, which was calculated by subtracting the total scale raw score for the IAAT form taken at the beginning of the semester from the total scale raw score for the IAAT form taken at the end of the semester. All students in the project completed Form A of the IAAT at the beginning of the study and Form B at the end. By conducting correlational analysis, we examined the relationships among these growth variables.

### Procedures

Project AAIMS research staff visited each class at the beginning of the course to present information about the study and gather informed consent. Students completed student assent forms during class and were given parent consent forms to take home. Teachers offered extra credit to students for returning signed consent forms (regardless of whether parents provided or withheld consent). The research staff also administered the Iowa Algebra Aptitude Test at the beginning and end of the study. Teacher ratings forms were distributed at the beginning (initial teacher rating of student proficiency) and the end (teacher rating of growth) of the study and collected by project staff.

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. Teachers administered probes according to a schedule with one-week intervals during which they

were to give two forms of one type of probe. Some teachers opted to give both probes on the same day; other teachers gave the two probes in on two different days. Some of the teachers were unable to administer probes as scheduled for the entire semester. The Xs shown in Table 4 indicate that the teacher administered the probe. As this table reveals, one teacher in District B was unable to administer the probes in Mid- April. In District C, one teacher was unable to administer the probes in End-April.

Table 4. Administration Schedule for Probe Forms by Period

Time/Period	Probe	District B		District C	
		T 1	T 2	T 3	T 4
Mid-Feb	A-33	X	X	X	X
Mid-Feb	A-34	X	X	X	X
End-Feb	E-33	X	X	X	X
End-Feb	E-34	X	X	X	X
Mid-Mar	A-35	X	X	X	X
Mid-Mar	A-36	X	X	X	X
End-Mar	E-35	X	X	X	X
End-Mar	E-36	X	X	X	X
Mid-Apr	A-31	X		X	X
Mid-Apr	A-32	X		X	X
End-Apr	E-31	X	X	X	
End-Apr	E-32	X	X	X	

Note. A-31 to A-36 denote *Basic Skills* probes; E-31 to E-36 denote *Content Analysis-Multiple Choice* probes

### Scoring Reliability

We hired and trained four pre-service teachers (subsequently referred to as “scorers”) to score the probes. The hiring process included a demonstration of correct scoring procedures for each type of probe and guided practice activities in which scorers worked with actual student papers. A final activity was the independent scoring of 10 student papers for each of the probe types. We used these probes to evaluate scoring reliability. 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. After training, the scorers’ mean interscorer agreement rates were 98.95% for the Basic Skills probes (range = 98.45% to 99.63%) and 95.84% for the Content Analysis-Multiple Choice probes (range = 94.27% to 96.90%). Scorers were informed that we would be checking their scoring accuracy levels throughout the project; they were able to earn bonus pay for maintaining high levels (>96% agreement) of accuracy in their scoring.

Following training, each scorer was assigned five classes with two forms of a probe per class to score (a total of 10 class sets of probes twice each month). Readers should note that the total of 20 classes includes additional algebra classes in Project AAIMS whose data are reported in other technical reports. Scorers also completed the data entry for the classes they were scoring. For each scorer, we conducted a scoring reliability check on two of the ten class sets in each scoring period (i.e., twice each month) by re-scoring all of the probes in those sets. The results of these interscorer reliability analyses are reported in the following section



## Results

### Scoring Reliability

Interscorer agreement rates revealed that scorers had high reliability on both types of probes. A total of 112 interscorer reliability checks were conducted across the four scorers throughout the 2005-2006 school year. The range of agreement for Basic Skills probes were between 98.1% and 100% with a mean of 99.1%. For Content Analysis-Multiple Choice probes, the interscorer agreement rates ranged from 94.5% to 100% with a mean of 98.7%.

### Descriptive Data on Score Ranges and Distributions

Table 5 lists the ranges, means, and standard deviations for each of the probes. For the *Basic Skills* probes, the number of correct answers was recorded. The total number of points possible for this probe was 60. 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 guess score represents the number of guess responses. The total possible correct and guess scores were 48 and 16, respectively.

Table 5. Descriptive Data for Algebra Probes Across Administration Sessions – Raw Scores

<b>Time Period</b>	<b>Probe</b>	<b>N</b>	<b>Score</b>	<b>Range</b>	<b>Mean</b>	<b>Standard Deviation</b>
<i>Algebra Basic Skills Probes</i>						
Mid-February	A-33	87	5 min.	4 – 36	13.55	6.66
	A-34	86	5 min.	4 – 38	14.23	6.71
Mid-March	A-35	75	5 min.	2 – 38	14.01	7.91
	A-36	57	5 min.	0 – 41	15.21	8.29
Mid-April	A-31	59	5 min.	5 – 37	16.64	6.34
	A-32	59	5 min.	4- 43	16.27	6.83
<i>Content Analysis-Multiple Choice</i>						
End-February	E-33	83	Correct	3-37	16.80	7.40
	E-33	83	Guess	0 – 7	1.70	1.50
	E-34	90	Correct	2 – 36	15.68	7.14
	E-34	90	Guess	0 – 12	1.96	2.11
End-March	E-35	81	Correct	4 – 41	21.12	9.12
	E-35	81	Guess	0 – 8	1.75	1.67
	E-36	80	Correct	6- 45	22.10	8.87
	E-36	80	Guess	0 – 8	1.38	1.17
End-April	E-31	59	Correct	9 – 45	26.00	8.87
	E-31	59	Guess	0 – 7	1.58	1.66
	E-32	60	Correct	9 – 45	28.00	9.30
	E-32	60	Guess	0 – 5	1.28	1.26

A close examination of Table 5 reveals two important points. First, the scores and correct responses on both types of probes increased as the semester went on. The increases were more dramatic for the Content Analysis-Multiple Choice probes than for the Basic Skills probes. This

finding suggests that the Content Analysis-Multiple Choice probes were more successful than Basic Skills probes in showing student growth over time. Second, the standard deviations were substantial (one-third to one-half the magnitude of the mean), suggesting that the measures would be beneficial in distributing students based on scores obtained on both probes. This finding is especially important if the probe data are to be used to identify students who are especially strong or weak in algebra.

We also examined whether scores obtained on the Basic Skills and Content Analysis-Multiple Choice probes differed by class type. As discussed earlier, students in three types of classes were participating in the study. Students in Algebra 1 were enrolled in a typical algebra course that was completed in a single semester (daily 90 minute periods on a block schedule). Students in Algebra 1A were completing the first half of Algebra 1 during the period of this study, while students in Algebra 1B were completing the second half of Algebra 1 during the study. We hypothesized that students in Algebra 1B would have the highest initial levels of performance (as they had already completed the first half of Algebra 1), but that the performance levels of Algebra 1 students would rise to similar levels by the end of the study. The means and standard deviations by class type are reported in Table 6 (Basic Skills) and Table 7 (Content Analysis-Multiple Choice).

Table 6. Descriptive Data for Basic Skills Probes by Class Type

<b>Time Period/Class Type</b>	<b>N</b>	<b>Range</b>	<b>Mean</b>	<b>Standard Deviation</b>
Mid-February				
<i>Algebra 1A</i>	14	6- 24	11.18	5.03
<i>Algebra 1B</i>	5	6- 17	10.80	4.39
<i>Algebra 1</i>	72	4-35	14.70	6.76
Mid-March				
<i>Algebra 1A</i>	16	6.5 – 23	13.28	4.99
<i>Algebra 1B</i>	5	2.5 – 17	10.50	5.72
<i>Algebra 1</i>	57	3- 38	15.29	8.46
Mid-April				
<i>Algebra 1A</i>	11	8.5 – 23	14.50	4.16
<i>Algebra 1B</i>	5	7 – 21.5	13.20	5.44
<i>Algebra 1</i>	44	5- 40	17.19	6.57

Table 7. Descriptive Data for Content Analysis-Multiple Choice Probes by Class Type

<b>Time Period/Class Type</b>	<b>N</b>	<b>Range</b>	<b>Mean</b>	<b>Standard Deviation</b>
End-February				
<i>Algebra 1A</i>	13	0 – 21	8.04	5.74
<i>Algebra 1B</i>	6	5 – 29	18.00	10.39
<i>Algebra 1</i>	73	3 – 34	15.15	6.25
End-March				
<i>Algebra 1A</i>	12	3 – 23.5	11.13	4.89
<i>Algebra 1B</i>	5	12.5 – 35.5	23.90	10.59
<i>Algebra 1</i>	68	2 – 43	21.55	8.45
End-April				
<i>Algebra 1A</i>	12	6- 25.5	12.29	5.69
<i>Algebra 1B</i>	5	24 - 33	27.70	3.49
<i>Algebra 1</i>	44	7 – 45	27.37	9.25

The data in Table 6 reveal consistent increases on the Basic Skills probes for students in Algebra 1 and Algebra 1A. Students in 1B had similar scores on the Basic Skills probes in the first two administrations, but showed an increase in the final testing period. In contrast to our hypothesis about the ordering of the class types, Algebra 1 students consistently achieved the highest scores on the Basic Skills probe, followed by Algebra 1A students and then the five students in Algebra 1B. Readers are reminded that our 1B sample included only five students, so it would be unwise to draw conclusions for this group.

Table 7 reveals that, as we had hypothesized, Algebra 1B students had the highest initial level of performance on the Content-Analysis-Multiple Choice probes (see End-February scores) and the performance levels of Algebra 1 students rose to similar levels by the end of the study (see End-April scores). Students in all three class types showed consistent increases over time on the Content Analysis-Multiple Choice measure.

#### Reliability of Probe Scores

The alternate form reliability of individual probes was evaluated by examining the correlation between two forms of a probe given during the same data collection session. We hypothesized that as the semester went on and students became more familiar with the probes, alternate form reliabilities would increase. We present the results of the alternate form reliability analyses in Table 8. These data suggest that the Basic Skills probes seemed to possess initially higher levels of reliability than the Content Analysis-Multiple Choice probes, but that these differences fade as students become more familiar with the measure. By the third round of data collection, the reliability levels for the two types of probes were essentially the same. With regard to our hypothesis, we observed increases in reliability for the Content Analysis-Multiple Choice probes over time, but not for the Basic Skills probes. The alternate form reliability coefficients obtained in this study are similar to those reported for the earlier study in Technical Report 12.

Table 8. Alternate Form Reliability Results for Single Probes

Time Period	Probes	Reliability
<i>Basic Skills</i>		
Mid-February	A-33 and A-34	.89
Mid-March	A-35 and A-36	.81
Mid-April	A-31 and A-32	.81
<i>Content Analysis-Multiple Choice</i>		
End-February	E-33 and E-34	.68
End-March	E-35 and E-36	.75
End-April	E-31 and E-32	.83

Note: All correlations were significant at  $p < .01$ .

We assessed the test-retest reliability of the probes by examining the correlation between the mean of two forms same forms of a probe administered across the two data collection time periods. For example, the two scores on the Basic Skills probes administered in mid-February were averaged and then correlated with the mean of the two scores on the Basic Skills probes administered in mid-March. Table 9 presents the results of the test-retest reliability analyses. We predicted that as the semester went on, test- retest reliability coefficients would increase. Our findings in Table 9 supported this hypothesis only for the Content Analysis-Multiple Choice probes. As with alternate form reliability, the coefficients we obtained in this replication study are similar in magnitude to those reported in Technical Report 12.

Table-9. Test-Retest Reliability Results for Aggregated Probes

Time Period	Reliability
<i>Basic Skills</i>	
Mid-February and Mid-March	.86
Mid- March and Mid-April	.79
<i>Content Analysis-Multiple Choice</i>	
End-February and End-March	.73
End-March and End-April	.78

Note: All correlations were significant at  $p < .01$ .

While these coefficients are lower than we would like, particularly for the Content Analysis-Multiple Choice measure, it is important to note that the test-retest period for these scores spanned four to six weeks. This duration is much longer than the five to seven day period often used to evaluate test-retest reliability. Our previous work with these measures and more typical test-retest time frames (see Technical Report 10) produced reliability estimates of .85 and .80 for the Basic Skills and Content Analysis-Multiple Choice measures, respectively.

### Concurrent Validity

The concurrent 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 teachers' evaluations of student proficiency and scores

and scores obtained from a norm-referenced test of algebra aptitude, the Iowa Algebra Aptitude Test (IAAT). We correlated students' scores on the February probes with the initial (spring) teacher rating of student proficiency and the form of the IAAT administered at the beginning of the study (labeled "Pre-IAAT" below). We also correlated students' scores at the end of the study (mid-April and end-April) with their scores on the IAAT administered at the end of the study ("Post-IAAT"). Correlation coefficients ranging from . to .58 (See Table 10) revealed that scores on the Basic Skills and the Content Analysis-Multiple Choice probes correlated with criterion measures at a low to moderate level. Student performance on the probes was more strongly correlated with their teachers' initial judgments of their proficiency in algebra at the beginning of the course than their scores on the Pre-IAAT scores. In comparing these results to the earlier study (Technical Report 12), we found a stronger relation between Basic Skills and the spring teacher rating (.58 here vs. .47 in the earlier study) and a comparable relation for the Content Analysis-Multiple Choice measure. Regarding relations with the IAAT, relations with Basic Skills were comparable, while the relation with Content Analysis-Multiple Choice was much weaker in this study than in the earlier investigation (.32 here vs. .57 earlier).

Table 10. Concurrent Validity Results for All Students

Time Period	Spring Teacher Rating	Pre-IAAT	Post-IAAT
<i>Basic Skills</i>			
Mid-February	.58	.51	
Mid-April			.50
<i>Content Analysis-Multiple Choice</i>			
End-February	.45	.42	
End-April			.49

Note: all correlations were significant at <.01 level.

We also examined concurrent validity to see if the magnitude of the correlation coefficients differed by class type. These results, reported in Table 11, reveal a large degree of variability in the correlation coefficients by class type. Spring teacher ratings were most highly correlated for Algebra 1B students, followed by students in Algebra 1. The correlations for both measures with spring teacher ratings were nonsignificant. When we examined relations between initial probe scores and the pre-IAAT, moderate relations were obtained with Basic Skills and low relations with Content Analysis-Multiple Choice for students in Algebra 1. No significant relations between the algebra probes and pre-IAAT were obtained for Algebra 1A students. Since Algebra 1B students took the IAAT exam in September instead of February, we were not able to look at the relationship between their February probe scores and Pre-IAAT scores. When we examined correlations between the final probe scores and the post-IAAT scores, we found moderate to strong relations with Basic Skills scores for Algebra 1 and Algebra 1A students, respectively. When the Content Analysis-Multiple Choice scores were correlated with the post-IAAT scores, a moderate relation was found for students in Algebra 1. All other correlation coefficients failed to achieve statistical significance. Again, these results should be interpreted with caution since we have a small number of students in Algebra 1A and Algebra 1B classes.

Table 11. Concurrent Validity by Class Type

Time Period /Class Type	Spring Teacher Rating			Pre-IAAT			Post-IAAT		
	N	<i>r</i>	<i>p</i>	N	<i>r</i>	<i>P</i>	N	<i>r</i>	<i>p</i>
<i>Basic Skills</i>									
Mid-February									
<i>Algebra 1A</i>	13	.38	.19	13	.26	.38			
<i>Algebra 1B</i>	5	.80	.10	-	-	-			
<i>Algebra 1</i>	67	.56	< .01	69	.51	< .01			
Mid-April									
<i>Algebra 1A</i>							8	.73	.03
<i>Algebra 1B</i>							4	.23	.77
<i>Algebra 1</i>							39	.45	< .01
<i>Content Analysis-Multiple Choice</i>									
End-February									
<i>Algebra 1A</i>	12	.22	.49	12	.21	.51			
<i>Algebra 1B</i>	6	.88	.02	-	-	-			
<i>Algebra 1</i>	68	.46	< .01	71	.29	.01			
End-April									
<i>Algebra 1A</i>							9	.13	.74
<i>Algebra 1B</i>							4	.52	.48
<i>Algebra 1</i>							41	.43	< .01

### Predictive Validity

We examined the predictive validity of the measures 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 Post IAAT scores, teacher rating of growth, algebra grade, and ITED scores. Readers should note that in District B, the ITED is administered in November, so the predictive relationship is opposite what would be expected (i.e., the ITED scores from November are predicting the spring probe scores). The results are reported in Table 12. As this table reveals, we found significant correlations between both types of probes and all of the criterion variables except for end-February probe and teacher growth rating. In general, predictive validity correlation coefficients were in the low to moderate range (.19 to .56).

In comparing these results to those obtained in the earlier study, the relations for Basic Skills are similar for post-IAAT and teacher growth ratings. The relations obtained with algebra grade and ITED scores are stronger than those obtained earlier, which were in the .33 - .36 range. For the Content Analysis-Multiple Choice measure, the post-IAAT, teacher growth rating, and algebra grade coefficients are smaller, while those for the ITED subtests are larger.

Table 12. Predictive Validity Results for All Students

Time Period	Post-IAAT	Teacher Growth Rating	Algebra Grade	ITED Comp	ITED Concept
<i>Basic Skills</i>					
Mid-February	.56*	.30*	.47*	.40*	.45*
<i>Content Analysis-Multiple Choice</i>					
End-February	.33*	.19	.37*	.42*	.30*

\* significant at <.01 level

We also examined predictive validity to see if the correlation coefficients differed by class type. These results are presented in Table 13. As with the earlier analyses by class type, readers are reminded to interpret the results with caution given the small sample sizes for Algebra 1A and Algebra 1B. For the post-IAAT scores, a similar pattern of results was obtained for both types of probes. Relations were very strong for students in Algebra 1A, with low to moderate relations for Algebra 1 students for the Content Analysis-Multiple Choice and Basic Skills measures, respectively. For teacher ratings of student growth completed at the end of the course, only one coefficient was significant (for Algebra 1 students on the Basic Facts) and it was in the low range. Coefficients in the low range were obtained for Algebra 1 students' grades with both probes, while a strong relation was obtained for Algebra 1B students' grades with the Content Analysis-Multiple Choice scores. When we examined relations with the ITED subtest scores, the coefficients were only significant for students in Algebra 1 and these were in the low to moderate range (.27 to .47). The magnitude of the relation between Basic Skills and the ITED Concepts subtest was particularly surprising. In past studies, the Basic Skills measure has tended to correlate most strongly with the Computation subtest; this was not true in the present data set.

Table 13. Predictive Validity Results by Class Type

Time Period /Class Type	Post-IAAT			Teacher Growth Rating			Algebra Grade			ITED Comp			ITED Concept		
	N	<i>r</i>	<i>p</i>	N	<i>r</i>	<i>p</i>	N	<i>r</i>	<i>p</i>	N	<i>r</i>	<i>p</i>	N	<i>r</i>	<i>p</i>
<i>Basic Skills</i>															
Mid-February															
<i>Algebra 1A</i>	6	.97	< .01	9	.33	.38	13	.39	.19	12	.18	.58	12	.02	.94
<i>Algebra 1B</i>	3	-.36	.76	4	.33	.67	5	.81	.09	5	.36	.55	5	.22	.71
<i>Algebra 1</i>	58	.52	< .01	66	.29	.02	70	.45	< .01	70	.39	< .01	70	.47	< .01
<i>Content Analysis-Multiple Choice</i>															
End-February															
<i>Algebra 1A</i>	5	.89	.04	8	.11	.80	11	.36	.27	10	.43	.22	10	.37	.30
<i>Algebra 1B</i>	4	.44	.55	5	.70	.18	6	.83	.04	6	.62	.19	6	.15	.77
<i>Algebra 1</i>	60	.29	.02	67	.20	.10	72	.35	< .01	71	.31	< .01	71	.27	.03



## Growth

As we considered our hypotheses regarding students' growth on the measures, we assumed that type of class would play an important role. Algebra 1B begin their course having already studied the content of the first half of Algebra 1. We assumed that these students would demonstrate the highest initial levels of performance, but that their rate of growth might not be as high as students in Algebra 1. Given our experiences in the Project AAIMS schools, students who enroll in the Algebra 1A/1B option typically prefer a slower paced course and are not as strong in mathematics as students who enroll in Algebra 1. Given this assumption, we expected that while Algebra 1 students might have lower initial levels of performance on the measures, their rates of growth would likely be steeper. Finally, we anticipated that students in Algebra 1A would have lower initial performance rates than either of the other two class types and would demonstrate a rate of growth similar to Algebra 1B students. To explore these hypotheses further, we conducted three types of analyses of student growth.

In our initial examination of student growth, we plotted the mean scores for students in each type of class every month (i.e., the data presented in Table 6) to see if the probes reflected growth similarly. Figure 1 displays the Basic Skills means for each probe administration for each of the three types of classes.

Figure 1. Monthly Growth Shown in Basic Skills Probes

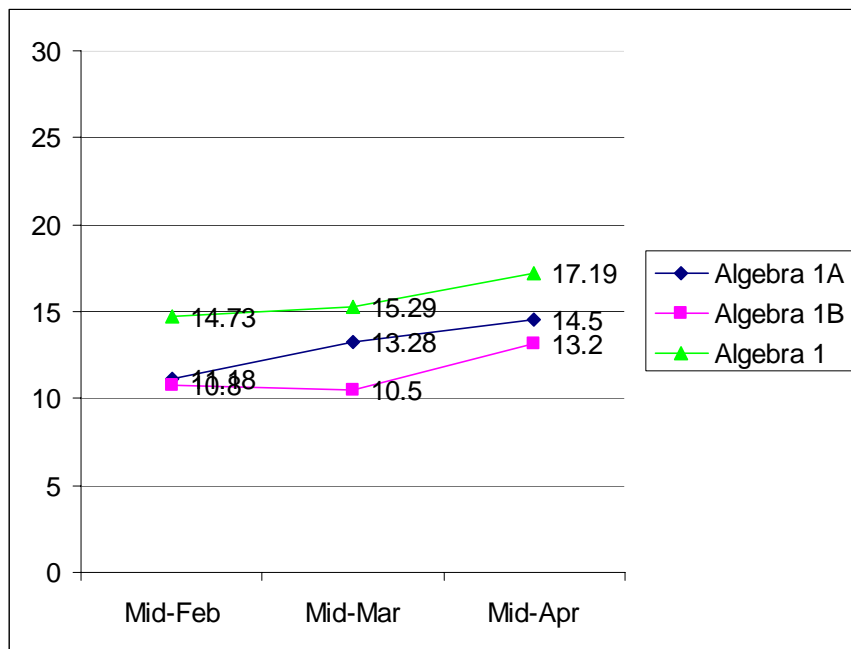


Figure 1 reveals that there was an apparent difference among class types in terms of showing growth. We were surprised to see that Algebra 1B students did not show a higher initial level of performance. As this figure reveals, there is a similar pattern of consistent increases for the Algebra 1 and Algebra 1A groups. In contrast, the Algebra 1B means showed an initial decline, followed by a remarkable increase. We next created a similar figure (Figure 2) displaying the results for students in each class type on the Content Analysis-Multiple Choice probes.

Figure 2. Monthly Growth Shown in Content Analysis-Multiple Choice Probes

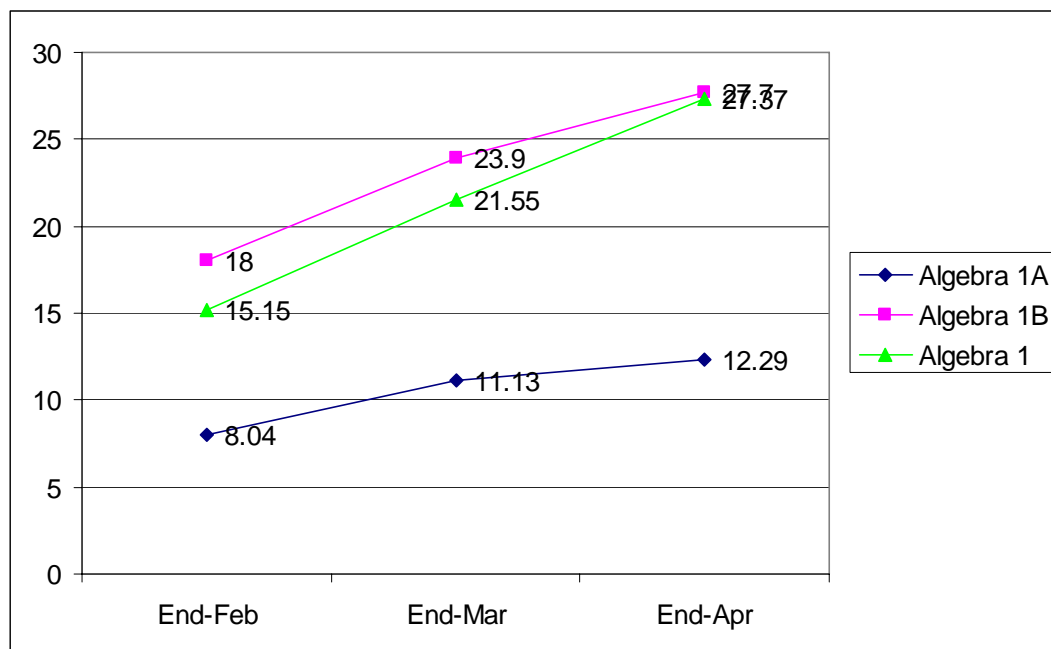


Figure 2 displays data that are quite consistent with our hypotheses. The figure reveals that the Algebra 1B group had a higher mean than Algebra 1 at the beginning of the study and that the Algebra 1 group’s mean score rose to the same level by the end of the study. Students in Algebra 1 appear to have the fastest rates of growth (steepest slope), followed by students in Algebra 1B. Algebra 1A students’ mean score was substantially lower than at the beginning of the study and they did not show as much growth as the other two groups did.

Our second set of analyses focused on individual students’ slope data, rather than the group means. As described earlier, students’ weekly growth was calculated using ordinary least square regression. We chose not to include students in the Algebra 1B class in this analysis because they participated in the study throughout the academic year. Their growth data is most appropriately reported using the full set of data and as a result, we opted to report their growth data in Technical Report 14, which addresses student growth across an entire school year. Therefore, this set of analyses was conducted only for Algebra 1 and Algebra 1A students. We examined slope values for all students (any students who had data for probes on at least two occasions) and “non-drop” students separately. Non-drop students are those who did not drop their algebra class prior to the end of the study. We were concerned that students who dropped the class might generate slopes that were calculated using just two data points and could potentially be misleading. Table 14 presents the range, mean, and standard deviation of slopes for Basic Skills and Content Analysis-Multiple Choice probes for all students and for non-drop students.

Table 14. Descriptive Statistics for Weekly Slope Values on Both Probes

	All Students				Non-Drop Students			
	N	Range	Mean	SD	N	Range	Mean	SD
Basic Skills	83	-1.87 – 2.75	.34	.90	78	- 1.87 – 2.63	.42	.83
Content Analysis-Multiple Choice	76	- 1.37 – 4.25	1.52	1.15	81	- 1.37 – 4.25	1.54	1.16

When we analyzed data for individual students’ slopes on each probe, we observed substantially larger weekly slope values for the Content Analysis-Multiple Choice measure in comparison to the Basic Skills measure. Students’ rates of improvement on the Content Analysis-Multiple Choice probes were three to four times greater than their rates of growth on the Basic skills probes. The data in Table 14 also reveal that removing students who dropped the class resulted in small increases in mean scores for both probes. It is important to note that the sample size changes more dramatically for Basic Skills than for Content Analysis-Multiple Choice probes when the analyses are limited to non-drop students. Table 14 also indicates that non-drop students grow .42 and 1.54 points each week on Basic Skills and Content Analysis-Multiple Choice probes, respectively. This result suggests that Content Analysis-Multiple Choice probes seem to be more sensitive to reflecting student growth.

We were also interested in determining whether the Basic Skills and Content Analysis-Multiple Choice probes reflected growth for each type of class similarly. Table 15 reports the average slope values on each of the measures by class type.

Table 15. Descriptive Statistics for Weekly Slope for Basic Skills and Content Analysis-Multiple Choice Probes by Class Type

	N	Range	Mean	SD
Basic Skills				
<i>Algebra 1A</i>	14	-1.63 – 2.75	.37	1.15
<i>Algebra 1</i>	71	-1.88 – 2.63	.31	.85
Content Analysis-Multiple Choice				
<i>Algebra 1A</i>	12	-.56 – 4.00	1.21	1.19
<i>Algebra 1</i>	66	-1.38 – 4.25	1.54	1.15

Table 15 reveals that students in Algebra 1A demonstrated slightly more growth on the Basic Skills probe than did those in Algebra 1 (.37 versus .31). However, we did not obtain similar results for Content Analysis-Multiple Choice. On this probe, students in Algebra 1 grew more than those in Algebra 1A (1.54 versus 1.21). While the pattern of results differed by class type for each of the probes, the Content Analysis-Multiple Choice measure would be the clear choice for progress monitoring for both class types given the substantially larger slopes for both classes in comparison to those obtained for Basic Skills.

It is important to note that all of the slope values for Content Analysis-Multiple Choice exceeded a weekly growth rate of .5. We have been using this benchmark as a goal in our

research. We anticipate that in order for algebra progress monitoring measures to be useful to teachers on a practical level, they must be able to expect to see scores grow by at least one point every two weeks (hence a weekly growth rate of .5). The results in Table 15 are encouraging that this benchmark is achievable for Content-Analysis-Multiple Choice probes, but less encouraging that the Basic Skills probes will meet this threshold. In the previous study, the results for both probes were less extreme. Mean weekly slopes on the Basic Skills measure ranged .32, .51, and .72, while mean weekly slopes for Content Analysis-Multiple Choice were .45, .60, and .86. The earlier results also favored the Content Analysis-Multiple Choice measure as the most sensitive, but the results were less dramatic than those obtained in the present study.

Finally, we were interested in determining whether the growth rates students obtained on the two types of probes were associated with other indicators of growth. As a result, we also examined the relationship of the slope values of both types of probes to teacher growth ratings and IAAT gain scores. These results are presented in Table 16. We found that there was a small but significant relationship between teacher growth ratings and slope values for the Content Analysis- Multiple Choice probes for all students and non-drop students. The same relation was not observed with Basic Skills slopes. Neither Basic Skills nor Content Analysis-Multiple Choice slopes were linearly related to IAAT gain scores. This result was surprising to us, as the IAAT had shown positive and relatively stronger relations to probe scores than other criterion measures we used. It may be that while the IAAT serves as an effective measure in dispersing students' scores at one point in time, it is not sensitive to changes in student performance and therefore does not function well when used to create gain scores.

Table 16. Correlations Between Slope Values and Teacher Growth Rating and IAAT Gain Scores

	All Students		Non-Drop Students	
	Teacher Growth Rating	IAAT Gain Score	Teacher Growth Rating	IAAT Gain Score
Basic Skills Slope	.08	- .06	.06	- .06
Content Analysis-Multiple Choice Slope	.27**	.06	.27**	.06

\*\* significant at < .05 level

### Summary and Future Research

The main purpose of this study was to replicate previous work examining the reliability, validity and sensitivity to growth of the Basic Skills and the Content-Analysis-Multiple Choice probes in two Iowa schools. One hundred five students in grades nine to twelve participated in the study. Data were gathered from February 2006 to April 2006. Over three months of data collection, students completed two Basic Skills probes and two Content Analysis-Multiple Choice probes each month.

We assessed the alternate form reliability and test-retest reliability of both types of probes. Our findings revealed that both types of probes possessed adequate levels of reliability with the Basic Skills probes demonstrating a higher initial level of reliability than the Content Analysis-Multiple Choice probes. By the final round of data collection, test-retest and alternate form reliability for both types of probes were comparable to each other and near or above the .80 level. We specifically hypothesized that as the semester went on and students became more familiar with the probes, both types of reliability would increase. Our findings supported this hypothesis for the Multiple Choice-Content Analysis probes, but not for the Basic Skills probes.

To assess the concurrent and predictive validity of both types of probes, we gathered data from a variety of indicators of students' proficiency in algebra including course grades, teachers' evaluations of student proficiency and growth, and performance on standardized assessment instruments including the ITED and IAAT exams. We assessed concurrent validity by examining the relationship of probe scores to teachers' evaluations of their students' proficiency and IAAT scores. We found that probe scores, in general, were moderately correlated with these criterion measures (.42 to .58), with stronger relations obtained for Basic Skills than for Content Analysis-Multiple Choice. When we disaggregated the data to examine concurrent validity by class type, we found that for Algebra 1 students both probes were significantly correlated with all of the criterion measures. In general, concurrent validity correlation coefficients were in the low to moderate range (.29 to .56) for Algebra 1 students. Correlations for students in Algebra 1A and 1B were more likely to deviate from the overall averages (both higher and lower) and generally failed to meet statistical significance criteria. Again, these results should be interpreted with caution since we have a small number of students in Algebra 1A and Algebra 1B classes.

We assessed predictive validity by investigating the relationship of the earliest probe scores (taken in the Mid-February and End-February) to teacher ratings of growth, students' end-term algebra grades, IAAT scores taken at the end of the semester, and ITED scores. Our results revealed that in general, both probes had low to moderate correlations (.30 - .56) with these indicators.

When examining the predictive validity of the probes using data disaggregated by class type, we found that the majority of the measures had low to moderate correlations with both types of probes for Algebra 1 students. The one exception was the relation between teacher growth ratings and the Content Analysis-Multiple Choice probes. For Algebra 1A students, the only significant relations were obtained between both types of probes and post-IAAT scores; these relationships were extremely strong (.97 for Basic Skills and .89 for Content Analysis-Multiple Choice). For Algebra 1B students, strong relations (.81, .83) were identified between both types of probes and algebra grades, however the coefficient for Basic Skills was not statistically significant. As with the earlier findings, this result suggests that the probes may function differently for students of varying algebra backgrounds and/or ability levels.

Finally, we were also interested in exploring whether the Basic Skills and Content Analysis-Multiple Choice probes reflected growth similarly for each type of class. To address

this issue, we conducted three sets of analyses. First, we examined graphs of the mean scores for students in each of the three class types for each data collection session. We observed a similar pattern of consistent increases on the Basic Skills measure for the Algebra 1 and Algebra 1B students. In contrast, the Algebra 1A students' means showed an initial decline, followed by small increases. On Content Analysis-Multiple Choice probes, it was found that there was an apparent difference among class types in terms of showing growth. Algebra 1 students had remarkably lower mean scores than other groups and did not show as much growth as the other groups did.

Next, we calculated individual student slope values for each type of probe and computed weekly rates of growth. We found that mean weekly slope values for students on Basic Skills and Content Analysis- Multiple Choice probes were .42 and 1.54 points, respectively. This result suggests that Content Analysis-Multiple Choice probes may be more sensitive in reflecting student growth.

Finally, we were also interested in whether the growth that students showed on both probes was related to teacher ratings of growth or IAAT gain scores. We found a small but significant correlation between teacher ratings of growth and Content Analysis-Multiple Choice slopes. No significant correlations existed between Basic Skills slopes and other indicators of growth.

Because this study was a replication of Technical Report 12, in this section we compare the results of the two studies in terms of alternate form and test-retest reliability, concurrent and predictive validity, and growth. To facilitate ease of communication in the following paragraphs, we use "Study 12" to refer to the initial study of student growth reported in Technical Report 12 and "Study 13" to refer to the present study.

In terms of alternate form reliability and test-retest reliability, both studies demonstrated that the two types of probes possessed adequate levels of reliability, with the Basic Skills probes demonstrating a higher level of reliability than the Content Analysis-Multiple Choice probes. In both studies we specifically hypothesized that as the semester went on and students became more familiar with the probes, both types of reliability would increase. Our results showed some differences between the two reports. In Study 12, we supported this hypothesis for both probe types with regard to alternate form reliability and for the Basic Skills probes with regard to test-retest reliability. In Study 13 our findings supported this hypothesis only for the Multiple Choice-Content Analysis probes.

Regarding concurrent validity, we found moderate correlations with other indicators of proficiency in both studies for both types of probes. We determined that the validity of the measures may vary for students in different levels of algebra and/or students with varying backgrounds in algebra. The predictive validity of both types of probes was supported in both studies by finding significant low to moderate correlations between probe scores and other indicators of proficiency. The only notable difference in Study 13 was that no significant correlation was found between the Content Analysis-Multiple Choice scores and teacher growth ratings.

In terms of how well both types of probes show student growth, in both studies we found that the Content Analysis-Multiple Choice probes were more sensitive to detecting student growth. This difference in mean weekly slope values was growth was more notable in Study 13 (1.54 versus .42) than in Study 12 (.61 versus .51).

The results of this study replicate previous work on the reliability and criterion validity of two algebra progress monitoring measures. In addition, we replicated positive findings regarding

the sensitivity of the probes to changes in student performance over time. Because this study and Study 12 were both conducted over a relatively short period of time (three months for Study 13, one semester for Study 12), future research should examine the characteristics of slopes obtained when progress monitoring is conducted across an entire school year. Future studies should also expand the range of probes investigated to include the Algebra Foundations measure, which was not examined here.

# APPENDIX

Standardized Administration Directions: Basic Skills

Standardized Administration Directions: Content Analysis-Multiple Choice

*Basic Skills* – Form 1

*Content Analysis-Multiple Choice* – Form 1

Teacher Rating of Student Proficiency

Teacher Rating of Student Growth





Project AAIMS 2005-06  
XXX High School

## PROBE STANDARD DIRECTIONS Basic Skills Probes

### GENERAL INTRODUCTION TO PROGRESS MONITORING:

The **FIRST** time you administer algebra probes, say:

*As you know, your class and other algebra classes at XXX High are working with Iowa State on a research project to learn more about improving algebra teaching and learning. Twice each month, we will be doing short algebra assessments, or probes, to monitor your learning in algebra. Remember that **all** students will be completing the probes and I will see the scores for all students, but your score will **only** be used for the Project AAIMS research if both you and your parent/guardian have given permission.*

*There are a few things you should know about these probes. First, you will be given a limited amount of time to work on the problems. These probes are different from classroom tests or quizzes and are not meant to be completely finished. What's important is that as you learn more about algebra in this class, your scores will improve. Second, keep in mind that the object of the probe is to correctly answer as many questions as you can in the amount of time given. There may be problems on the probes that are difficult or unfamiliar. Please look at each problem. If you do not know how to answer it, skip it, and go on 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 skipped and try to solve them. Third, your scores on these probes will be used to see your progress in algebra. Because of this, it's important that you try your best. Do you have any questions at this point?*

### BASIC SKILLS PROBES:

Hand out probe A-31 (with the sample page), keeping the probes face down. Ask students to keep the probes face down and write their name and the date on the back of the probe.

Give the standard directions:

The **FIRST** time you administer BASIC SKILLS algebra probes, say:

*Please turn your paper over. This sample page shows some examples of the types of problems on the Basic Skills probes. The questions include solving algebra equations using basic math facts, simplifying expressions by combining like terms, using the distributive property to simplify expressions, and solving proportion, or ratio problems. Now we'll take a minute so you can practice doing a Basic Skills probe. If you finish before I say 'Stop', please do NOT turn to the next page. Any questions? Ready, begin. [Time for 1 minute] Stop, pencils down.*

*Now that you've had a chance to try out this type of probe, do you have any questions? [Only answer procedural questions—do not suggest ways to solve the problems.]*

*Now we'll do the first Basic Skills probe. You will have 5 minutes to work on this two-page probe. Remember, your job is to answer as many problems correctly as you can in 5 minutes. Please look at each problem, but if you do not know how to do it, skip it and move on. If you get to the end of the probe before the time is up, go back and work on the more difficult problems. When you solve the simplifying questions, be sure to go as far as you can with your answer.*

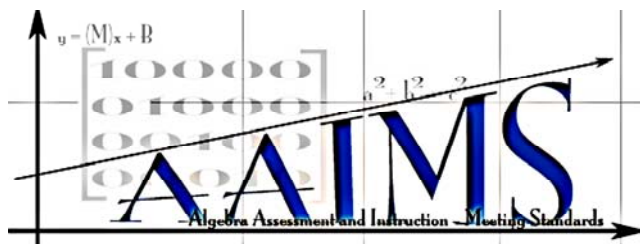
*When I say begin, please turn past the sample page and begin working. You will have 5 minutes. Please do your best work*

Time for 5 minutes. When the timer goes off, say *Stop*. Please put your pencils down, and collect student papers.

For **ALL OTHER** administrations, hand out the probes face down and say

*Please write your name and the date on the back of your paper. You are going to do a Basic Skills probe. You will have 5 minutes to work. Remember to try and complete as many problems correctly as you can in the time allowed. When you are simplifying, be sure to go as far as you can with your answer [write your answer in lowest terms]. Please do your best work. Ready? Begin.*

Time for 5 minutes. When the timer goes off, say *Stop*. Please put your pencils down, and collect student papers.



## PROBE STANDARD DIRECTIONS Textbook Probes

### E (TEXTBOOK) PROBES:

Hand out probe E-31 (with the sample page), keeping them face down. Ask students to keep the probes face down and write their name and the date on the back of the probe.

Give the standard directions:

The **FIRST** time you administer the TEXTBOOK algebra probes, say:

*The problems on this probe come from the chapters of the book, but they are not in any special order. For example, a problem from Chapter 1 could be the last problem on the probe. Please look at each problem and decide if you know how to do it. If you do, go ahead and solve the problem. If aren't certain or think you can't solve the problem, skip it and move to the next one. Don't spend too much time on any one problem. The object of the probe is to answer as many problems correctly in the time available. Once you get to the end, go back and work on the difficult problems. Remember that you may earn partial credit by showing your work even if you can't solve the entire probe. Do NOT make wild guesses because this will cause you to lose points on the probe.*

*Please turn your paper over. This sample page shows some examples of the types of problems on the Textbook probes. The problems on this probe are drawn from the different types of problems you are learning in the textbook. The questions are multiple choice. Each problem is worth 3 points, but you can earn partial credit by showing your work. Unless you are completely certain of the correct answer, the best strategy is to show your work. If you do not know the answer, you should NOT make wild guesses. You will lose points from your total score on the probe when you make wild guesses.*

*Look at the three boxes in the first row labeled A, B, and C. You'll notice that all three have answers and that the problem is the same for all three. Look at the box for Student A. She thought she knew the correct answer, so she just circled her choice at the bottom. Unfortunately, she was incorrect, so she will lose*

a point for this problem. Student B showed his work, but did not know how to finish the problem. Because he did part of the problem correctly, Student B will earn 1 out of 3 points on this problem. Student C solved the problem, but made an error, so her final answer is not correct. Because she showed her work, she will earn 1 out of 3 points on the problem for the part she has done correctly. As you can see from these examples, it is important to show your work on these probes.

Let's take a minute so you can practice doing a Textbook probe. If you finish before I say 'Stop', please do NOT turn to the next page. Any questions? Ready, begin. [Time for 1 minute] Stop, pencils down.

Now that you've had a chance to try out this type of probe, do you have any questions? [Only answer procedural questions—do not suggest ways to solve the problems.]

Now we'll do the first Textbook probe. You will have 7 minutes to work on this two-page probe. Remember, your job is to answer as many problems correctly as you can in 7 minutes. Please look at each problem. If you do not know how to do it, skip it and move on. If you get to the end of the probe before the time is up, go back and work on the problems you skipped. Remember that you may earn partial credit by showing your work even if you can't solve the entire problem. Do NOT make wild guesses because this will cause you to lose points on the probe.

When I say begin, please turn past the sample page and begin working. You will have 7 minutes. Please do your best work. Ready? Begin.

Time for 7 minutes. When the timer goes off, say Stop. Please put your pencils down, and collect student papers.

For **ALL OTHER** administrations, hand out the probes face down and say

*Please write your name and the date on the back of your paper. You are going to do a Textbook probe. You will have 7 minutes to work. Remember to try and complete as many problems correctly as you can in the time allowed. If you're not sure of an answer, skip the problem and come back to it if you have time left. Remember that you can earn partial credit by showing your work. Do NOT make wild guesses. Please do your best work. Ready? Begin.*

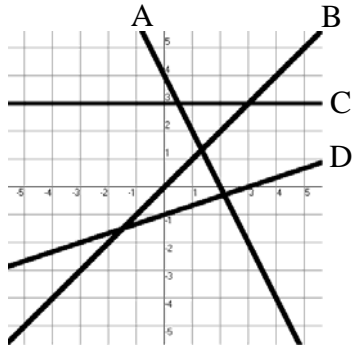
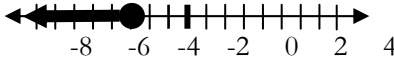
Time for 7 minutes. When the timer goes off, say Stop. Please put your pencils down, and collect student papers.

Solve: $9 + a = 15$ $a =$
Evaluate: $12 + (-8) + 3$
Simplify: $2x + 4 + 3x + 5$
Solve: $12 - e = 4$ $e =$
Simplify: $4(3 + s) - 7$
Simplify: $b + b + 2b$
Solve: $\frac{b}{6} = \frac{12}{18}$ $b =$
Simplify: $7 - 3(f - 2)$
Evaluate: $-5 + (-4) - 1$
Solve: $63 \div c = 9$ $c =$
Simplify: $2(s - 1) + 4 + 5s$
Simplify: $8m - 9(m + 2)$
Solve: 3 feet = 1 yard _____ feet = 9 yards
Evaluate: $4 - (-2) + 8$
Simplify: $2k + 3 - 5(k + 7)$

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

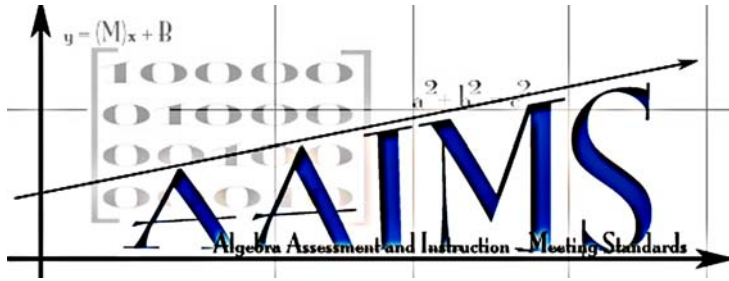
Solve: $3 \cdot 8 = m$ $m =$
Evaluate: $-9 + 5 + 8$
Simplify: $x + 2(x - 5) - 3$
Solve: $d - 5 = 4$ $d =$
Simplify: $5(3 + f) - 2f + 6$
Simplify: $5 - 2b + 4(b + 3)$
Solve: 4 quart = 1 gallon _____ quarts = $3 \frac{1}{4}$ gallons
Simplify: $4(y + 1) - 8y$
Evaluate: $14 - 7 + (-3)$
Solve: $\frac{36}{6}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

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

<p>Solve: <math>3x + 4 = 19</math></p> <p>a) <math>x = 8</math> b) <math>x = 22</math> c) <math>x = 15</math> d) <math>x = 5</math></p>	<p>Evaluate <math>a^2 - b \div 2</math> when <math>a = 4</math> and <math>b = 6</math></p> <p>a) 1 b) 5 c) 10 d) 13</p>	<p>Which line on the graph is <math>y + 2x = 4</math> ?</p>  <p>a) Line A b) Line B c) Line C d) Line D</p>	<p>Simplify: <math>3(m + 2) + 2(m - 1)</math></p> <p>a) <math>5m + 4</math> b) <math>5m + 1</math> c) <math>6m + 8</math> d) <math>6m - 8</math></p>
<p>Evaluate the expression: <math>4^{-2}</math></p> <p>a) <math>-16</math> b) <math>\frac{1}{16}</math> c) <math>\frac{1}{8}</math> d) <math>-8</math></p>	<p>Solve the linear system: <math>x - y = 4</math> <math>x + 2y = 19</math></p> <p>a) <math>(-1, -5)</math> b) <math>(5, 8)</math> c) <math>(-2, 19)</math> d) <math>(9, 5)</math></p>	<p>This graph shows the solution for which equation?</p>  <p>a) <math>x &gt; -3</math> b) <math>2x \leq -6</math> c) <math>-3x &gt; 9</math> d) <math>3x \geq 9</math></p>	<p>Write the equation in slope-intercept form if <math>m = \frac{1}{2}</math> and <math>b = 3</math></p> <p>a) <math>y = 2x + 3</math> b) <math>y = 3x + \frac{1}{2}</math> c) <math>x = \frac{1}{2}y - 3</math> d) <math>y = \frac{1}{2}x + 3</math></p>

<p>Evaluate <math>d + 3e^2</math> when <math>d = 5</math> and <math>e = 2</math></p> <p>a) 11 b) 23 c) 17 d) 10</p>	<p>Solve: <math>6c + 4 = -3c - 14</math></p> <p>a) <math>-\frac{10}{3}</math> b) <math>-2</math> c) 2 d) 6</p>	<p>Find the slope of a line through <math>(1, -1)</math> <math>(5, 2)</math></p> <p>a) <math>\frac{1}{5}</math>      b) <math>\frac{3}{4}</math> c) <math>-6</math>      d) <math>-\frac{4}{3}</math></p>	<p>Simplify: <math>6(2b - 3) - 3(2 - b)</math></p> <p>a) <math>15b - 24</math> b) <math>9b - 9</math> c) <math>9b + 12</math> d) <math>15b + 12</math></p>
<p>Simplify the expression:</p> $\frac{a^2}{ab^3} \cdot \frac{b^4}{a^3}$ <p>a) <math>\frac{a^8}{a^3b^3}</math>      b) <math>\frac{ab^8}{a^4b^3}</math> c) <math>\frac{b}{a^2}</math>      d) <math>\frac{b}{a}</math></p>	<p>Solve the linear system: <math>-6x + 3y = -6</math> <math>2x + 6y = 30</math></p> <p>a) (6, 3) b) (3, 4) c) (2, 6) d) (4, -3)</p>	<p>Simplify <math>b^2 - 4b + 2b^2 + 7 - 5</math></p> <p>a) <math>3b^2 - 4b + 2</math> b) <math>2b + 2</math> c) <math>-b^2 - 4b + 12</math> d) <math>3b^2 - 4b + 12</math></p>	<p>Write the equation of a line through <math>(5, 3)</math> <math>(4, 9)</math>. Use point-slope form.</p> <p>a) <math>y + 1 = 2(x - 4)</math> b) <math>y + 4 = -6(x - 1)</math> c) <math>y - 3 = -6(x - 5)</math> d) <math>y = -6x + 30</math></p>





Teacher \_\_\_\_\_

**Project AAIMS:  
Algebra Assessment and Instruction:  
Meeting Standards**  
XXX High School  
Spring 2006

Directions: For each student, rate his or her **general proficiency in algebra** relative to other students in your class(es). A rating of "1" indicates a very low level of proficiency, "4" indicates average proficiency, and "7" indicates exceptional proficiency. Try to spread student ratings across the **full range of the scale**, not clustering students only in the middle or toward one end.

<u>Student</u>	<u>Algebra Proficiency</u>						
	Low	Average					High
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7

Teacher \_\_\_\_\_

**Project AAIMS:  
Algebra Assessment and Instruction:  
Meeting Standards  
XXX Senior High**

### Teacher Rating of Student Progress

Directions: Below is a list of the students you teach. Please rate the amount of progress or improvement each student has made in algebra during this course. A rating of "1" indicates no growth or a decrease in level of performance, "3" indicates average progress, and "5" indicates exceptional progress, far beyond what you expected. If the list includes a student who has dropped the class, just draw a line through his/her name. Thank you!

<u>Student</u>	<u>Algebra Progress</u>				
	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