TECHNICAL REPORT #15:

Reliability, Criterion Validity, and Changes in Performance Across Three Points in Time: Exploring Progress Monitoring Measures for Middle School Mathematics

Anne Foegen, Karly Klein, Linda Lind, and Cynthia L. Jiban

RIPM Year 3: 2005 – 2006

Date of Study: October 2005 – May 2006

August 2008

The College of Education & Human Development

UNIVERSITY OF MINNESOTA

IDEAs that Work

U.S. Office of Special Education Programs

Produced by the Research Institute on Progress Monitoring (RIPM) (Grant # H324H30003) awarded to the Institute on Community Integration (UCEDD) in collaboration with the Department of Educational Psychology, College of Education and Human Development, at the University of Minnesota, by the Office of Special Education Programs. See progressmonitoring.net.

Abstract

The present study examines reliability, criterion validity, and growth data for 563 middle school students in Grades 6, 7, and 8 in two Midwestern middle schools. Two to three forms of six different mathematics progress monitoring measures were administered in the fall, winter and spring of the 2005-06 school year. In the fall, two rounds of data were collected to evaluate test-retest reliability. The measures included in the study were Monitoring Basic Skills Progress-Computation, Monitoring Basic Skills Progress-Concepts and Applications, Estimation, Basic Facts, Complex Quantity Discrimination, and Missing Number. Criterion measures included teacher ratings of student proficiency, course grades, and scores on norm-referenced tests of mathematics and reading achievement. Reliability results (both alternate-form and test-retest) were adequate for most measures; using the mean of two forms provided the most reliable estimates of student performance. Criterion validity (concurrent and predictive) coefficients reflected a wide range in the strength of relations between the mathematics progress monitoring measures and the external criteria. The measures with the strongest criterion validity at each grade level were the Monitoring Basic Skills Progress-Concepts and Applications and Monitoring Basic Skills Progress-Computation measures at Grade 6, the Monitoring Basic Skills Progress-Concepts and Applications measure at Grade 7, and the Complex Quantity Discrimination measure at Grade 8. These same measures were among those reflecting the most rapid rates of improvement at each grade level.

Reliability, Criterion Validity, and Changes in Performance Across Three Points in Time: Exploring Progress Monitoring Measures for Middle School Mathematics

Background Information

Limited options are available for middle school teachers seeking progress monitoring measures in mathematics. Foegen and her colleagues (Foegen, 2000, Foegen & Deno, 2001) have conducted initial research on potential measures involving estimation and basic fact proficiency. Helwig and his colleagues (Helwig, Anderson, & Tindal, 2002; Helwig & Tindal, 2002) have examined the use of a concept-based measure to predict performance on a high stakes state assessment. Unfortunately, Helwig and Tindal found this concept-based measure did not prove to be sensitive to changes in performance at a level that would be useful to practitioners was observed across the school year, as the mean total scores changed less than 2 points over the course of an academic year.

Foegen (2000) and Foegen and Deno (2001) examined the technical characteristics of an estimation measure in which students chose the best estimate from among three options, each of which differed by a factor of ten. The problems included both computational estimation (i.e., 219 + 876 is about [10 100 1,000]) and story-based, or contextual problems (i.e., Joel earns $4.25 per hour stacking shelves at the grocery store. About how much will he earn in 20 hours? [$0.80 $8 $80]). Foegen's (2000) sample included 105 sixth grade students at a Midwestern middle school. The problems included all four operations and whole numbers, fractions, decimals, and percents. The answer alternatives differed by a factor of 10 to encourage students to use estimation and number sense or mental math, rather than exact computation, to select the most appropriate estimate. Foegen and Deno's (2001) sample included 100 students in grades six to eight in an urban middle school. Research results for these studies have documented internal consistency, alternate form, and test-retest reliability coefficients ranging from r = .67 to .93. Criterion validity coefficients with outcome variables such as grades, teacher ratings, and standardized test scores ranged from r = .29 to .62, with most in the r = .40 to .50 range. When weekly rates of growth on the estimation measures were examined, students increased .25 points per week on the estimation measures over a 10-week period.

In the same studies, Foegen and her colleagues also investigated the use of a simple basic facts task that required students to solve single digit combinations (0-9) in each operation. Reliability estimates for the basic facts measure (internal consistency, alternate form, and test-retest) ranged from r = .79 to .95. Criterion validity coefficients (with the same outcome measures as for estimation) ranged from r = .44 to .66, with most in the r = .40 to .50 range. Students' weekly growth on the basic facts measure was more than double their growth on the estimation measure, with a mean increase of .55 problems per week. The results of Foegen's work suggested that both measures have acceptable levels of reliability and validity, but that the facts measure is more likely to be sensitive to small changes in student performance.

In the absence of widely available measures specifically designed for secondary mathematics, many teachers and educational agencies are currently using two measures developed to reflect elementary mathematics curriculum as a means to monitor the progress of secondary students with disabilities who are performing below grade level in mathematics. Lynn Fuchs and her colleagues at Vanderbilt University developed the Monitoring Basic Skills Progress-Computation measure (Fuchs, Hamlett, & Fuchs, 1998) and the Monitoring Basic Skills Progress-Concepts and Applications measure (Fuchs, Hamlett, & Fuchs, 1999). The measures are distributed commercially in combination with a computerized system that provides graphing, error analyses, and instructional recommendations. The measures have documented levels of technical adequacy and have been examined in several published studies.

Teachers of secondary students with disabilities often use these measures by first determining the grade level at which a middle school or high school student is functioning; they then use the corresponding level of MBSP measure to monitor the student’s progress. In Iowa, the Heartland Area Education Agency has published normative data on the measures from grades K through 12 (Heartland AEA 11, n.d.). Special education teachers within the area education agency use these data to compare a student’s performance to grade level norms and to set annual IEP goals.

We were unable to locate any research evidence in the literature addressing the efficacy of using mathematics measures designed for elementary students with middle school and high school students. Although the normative examples in the field provide data against which to compare student performance, we have not been able to identify any published accounts in which criterion validity data are reported that demonstrate the degree of relationship between students’ scores on the elementary mathematics measures and performance on external criterion measures, such as a state or district-wide assessment. Technical adequacy data of this type are increasingly important as IDEA 2004 reforms, in combination with NCLB provisions for Highly Qualified Teachers and access to grade level curriculum are increasing the numbers of students with disabilities who are participating in grade level mathematics curriculum, often in general education settings. One might argue that the use of progress monitoring measures that reflect the student’s functional level of performance may be valid when the student is receiving an individually designed mathematics instructional program that corresponds to that level. However, it is less clear that that the measures will have comparable levels of technical adequacy when the student in participating in grade level curricula that may deviate from the content of the elementary measures.

Purpose of the Study

The purpose of this study was to examine the technical adequacy of six potential measures for monitoring student progress in mathematics at the middle school level. The measures examined included two measures that have been investigated previously with middle school students (facts, estimation), two measures that have documented technical adequacy for use at elementary grade levels (MBSP Computation, MBSP Concepts and Applications), and two measures that represent extensions of elementary numeracy measures developed through the Research Institute on Progress Monitoring's (RIPM) research efforts. This study was designed to explore the reliability (alternate form and test-retest) and criterion validity of the measures. In addition, the measures were administered in the fall, winter, and spring of an academic year to provide preliminary data about the measures' sensitivity to changes in student performance over time.

Research Questions

The following research questions guided our study:

1. What is the reliability of the middle school mathematics measures?
2. What is the criterion validity of the measures?
3. How much academic growth do students display on each of the measures over the course of one academic school year?

Method

Setting and Participants

The study was conducted two Midwestern school districts. The first district (A), in a suburban area, was comprised of six schools: four of them grades K-5, one grades 7-8, one grades 8-9, and one grades 10-12. During the 2005-06 school year, District A enrolled 4,577 students, with 51 percent being male, 91 percent white, 3 percent African American, 3 percent Asian, 3 percent Hispanic, and less than 1 percent other ethnicities. Ten percent of the students qualified for free or reduced lunch, 8.8 percent of students were receiving special education services, and 2.5 percent were Limited English Proficient. The second district (B) was located on the fringe of a Midwestern town. District B had three schools: two grades K-6 and one grades 6-12. During the 2005-06 school year, District B enrolled 1,433 students, with 51 percent being male, 98 percent white, 1 percent Hispanic, and 1 percent other ethnicities. Eighteen percent of the students in District B qualified for free or reduced lunch, eleven percent were receiving special education services, and none were Limited English Proficient.

Teacher participants in the study included three grade 6 general education teachers, three Grade 7 general education teachers, one Grade 8 general education teacher, and one Grade 7 special education teacher. Each of the general education teachers taught 4 or 5 sections of mathematics. The Grade 8 teacher and her students were from District B; the remainder of the teachers were from District A.

Five hundred sixty three students participated in the study (242, 264, and 57 students in Grades six, seven, and eight, respectively). Table 1 displays the demographic characteristics of the student participants.

Table 1

Demographic Characteristics of Student Participants

	6th Grade (n = 242)	7th Grade (n = 264)	8th Grade (n = 57)	Total Percentage
Gender				
Male	114	110	24	44
Female	120	145	33	53
Ethnicity				
White	215	239	57	90.7
African American	8	5	0	2.3
Asian	6	7	0	2.3
Other ethnicities	5	4	0	1.6
Free/reduced lunch	14	16	13	7.6
ESL Services	2	1	1	.07
Special Education Services	10	18	5	5.9

Measures

Middle school mathematics progress monitoring measures. Six mathematics progress monitoring measures were investigated in this study: Monitoring Basic Skills Progress Computation (MBSP-Comp; Fuchs, Hamlett, & Fuchs, 1998), MBSP Concepts and Applications (MBSP-ConApp; Fuchs, Hamlett, & Fuchs, 1999), Basic Facts, Estimation, Complex Quantity Discrimination, and Missing Number. The MBSP-Comp task reflected 6th grade curriculum content and required students to compute addition, subtraction, multiplication, and division problems involving whole numbers, fractions, and decimals. The task consisted of 25 problems; students had six minutes to answer as many problems as possible. We scored this measure by counting the number of correct digits in each answer and summing to get a total score.

The MBSP-ConApp task reflected Grade 6 computation curricula and required students to answer problems related to numeration, applied computation, measurement, geometry, percentages, charts and graphs, word problems, ratios and probability, proportions, and variables. Students had seven minutes to complete 24 items. We scored this measure by counting the number of correct responses.

The Basic Facts task required students to solve single digit computation combinations (0-9) in each of the four operations. The task consisted of 80 items; students had one minute to respond. We scored the probe by counting the number of correct responses.

The Estimation probe required students to select the most appropriate estimate from three alternatives when presented with a problem involving either computation (i.e., 1.8 x 4 is about) or contextual (story) problems (i.e., Jared had $15.50 from mowing lawns. He made another $9.25 from working at the retirement center. About how much does Jared have now?). The problems included all four operations and whole numbers, fractions, decimals, and percents. The answer alternatives differed by a factor of 10 (i.e., 4, 40, 400) to encourage students to use estimation and number sense or mental math, rather than exact computation, to select the most appropriate estimate. The task consisted of 40 problems and students had 3 minutes to circle the correct alternative for each problem. We scored this measure by counting the number of correct responses and then subtracting from this value one-half the number of incorrect responses. This scoring method provides a correction for the potential that students will guess in responding to the multiple-choice item format and has been found in previous research to produce scores with higher levels of technical adequacy (Foegen, 2000).

The Complex Quantity Discrimination task consisted of 44 pairs of numbers or equations. The task required students to analyze the pairs of quantities (i.e., 40 + 7 and 70 + 4) and write the appropriate symbol (greater than (>), less than (<), or equal to (=)) in a box between the quantities. Students had one minute to complete the task. We scored the probe by counting the number of correct responses.

To complete the Missing Number task, students were presented with a series of three numbers and one blank indicating a missing fourth element in the sequence (the position of the blank varied). The number patterns included counting sequences, multiples (i.e., 3, 6, 9, ___), multiples starting on an off digit (i.e., 11, 14, ___, 20) and halving/doubling patterns. Students responded by writing the missing number in the blank. The task consisted of 44 items; students had one minute to respond. We scored the probe by counting the number of correct responses. Samples of the measures, including administration directions, are provided in Appendix A.

Criterion measures. The criterion measures used in the study included teachers' ratings of their students' overall math proficiency, math grades, and standardized test scores. Teachers were asked to rate each student's general proficiency in math relative to other students in his/her class, on a Likert scale from one to seven, with one being least proficient and seven being most proficient. Teachers were asked to spread student ratings across the full range of the scale, not clustering students' ratings in the middle or toward one end. All teachers completed student ratings in the fall and spring, concurrent with the respective probe administration procedures. Prior to any statistical analyses, rating scores were standardized for each teacher by converting the scores to *z* scores. Teacher ratings were available for all students. A sample of the teacher rating form is provided in Appendix B.

Data on student grades were collected by converting each student's end of year letter grade in mathematics to a number on a 4.0 scale. For example, a grade of A was recorded as 4.0, A- as 3.67, B+ as 3.33, B as 3.0 and so forth. Grade data were only available for 8th grade students in district B.

Two different norm-referenced achievement measures were used in the study: the Iowa Tests of Basic Skills (ITBS; H.D. Hoover, S.B. Dunbar, D.A. Frisbie, 2000) and the Northwest Achievement Levels Test (NALT; Northwest Evaluation Association, 2002). The ITBS is designed to be a valid and reliable measure that can be used to describe students' developmental level within a content area. The test was designed to measure students' critical thinking skills (i.e., interpretation, inferential thinking, classification, analysis, comparisons). The ITBS is designed for students in Kindergarten through eighth grade. It takes approximately five and one half hours to complete the entire battery of tests. Internal consistency reliability, based on the performance of a standardization sample from 2000, ranged from r = .66 to .93 (Salvia & Ysseldyke, 2007). The authors have provided evidence for the content validity of the measure in three ways. First, they consulted curriculum guides, textbooks, classroom teachers, and school administrators when writing test items. Second, test items were administered to a national sample of students at each grade level. Third, each test item was reviewed for content fit and item bias by a group of experts.

The Math Problem Solving and Data Interpretation, and the Math Concepts and Estimation Subscales, as well as the Math Total Score were used in our analyses. The Math Concepts test includes numeration and number systems, whole numbers, geometry, measurement, fractions, money, and number systems problems. The Math Problem Solving test is comprised of two sections. One section requires students to solve short word problems while the other section requires students to interpret information in graph and table format. We recorded the national percentile rank earned by each student on each scale for use in our analyses.

The NALT is a nationally normed, standardized achievement test of reading and mathematics. The mathematics portion of the NALT measures student proficiency in computation; number sense; shapes, space and measurement; patterns, relations and functions; randomness, uncertainty and discrete mathematics; and data investigation. Marginal reliability estimates for the mathematics test at grades six through eight, as reported in the test manual (Northwest Evaluation Association, 2003), ranged from r = .94 to .96. Test-retest reliability coefficients for these grade levels in mathematics ranged from r = .89 to .94. Correlation coefficients between the NALT and the Stanford Achievement Test-9 and the Iowa Tests of Basic Skills ranged from r = .84 to .88. Typically, obtained RIT scores fall between 150 and 300. In contrast to percentile rank scores, RIT scores are equal interval. Students completed the NALT in the fall and again in the spring; both mathematics scores were recorded for use in the analyses.

Procedures

Project staff visited each classroom to explain the study to students and describe the assent/consent process. Student assent forms were completed in class; parent consent letters were sent home with students. Teachers collected the consent forms as the students returned them.

All students present in each class participated in four rounds of data collection spread across the school year (fall, winter, spring). The first two rounds of data were collected in mid October. Data were collected again in mid February and in mid April. Group administration of the tasks took place during normal class periods, with each data collection session lasting between 30 and 40 minutes. No makeup sessions were held for students who were absent on a day when data were collected. For each of the probe types, the administration of the multiple forms was preceded by a brief introduction to the measure and a practice exercise in which students completed several problems similar to those on the measure to insure they were familiar with the task demands.

The measures were subdivided into four groups to minimize the demands for testing time and assigned on the basis of teacher. Group A included two forms of MBSP-Comp and two forms of Estimation. Group B included two forms of MBSP-ConApp and two forms of Basic Facts. Group C was subdivided into two groups, which differed with regard to one of the three measures administered. The C-Estimation Group included three forms of Complex Quantity Discrimination, three forms of Missing Number, and two forms of Estimation. The C-Facts Group included three forms of Complex Quantity Discrimination, three forms of Missing Number, and two forms of Basic Facts. The order of the tasks within each group was counterbalanced across classes to control for any possible order effects. Table 2 shows the assignment of measures by teachers.

Table 2

Assignment of Measures by Teacher

		Period							
Teacher	Grade	1	2	3	4	5	6	7	8
Gen Ed 1	6	C-Fact	C-Fact		C-Est	C-Est			
Gen Ed 2	6		A	A	A	A			
Gen Ed 3	6	B	B		B	B			
Gen Ed 4	7	A	A				A	A	
Gen Ed 5	7	C-Est	C-Est					C-Fact	C-Fact
Gen Ed 6	7	B	B					B	B
Spec Ed 1	7		C-Est						
Gen Ed 7	8		C-Est	C-Est	C-Est		C-Est	C-Est	

The NALT and ITBS were group administered to all students by district staff using their district's standardized assessment procedures, but only those for whom parental consent and student assent had been obtained are reported in this paper. Standard administration directions (as specified in the teacher directions) and timing requirements were used.

Trained data collectors gathered all the data. Each data collector participated in a small-group training session lasting approximately one hour. Training was provided by the graduate research assistant using a PowerPoint presentation and assessment materials used in this study.

Project staff completed all of the scoring and data entry. At least ten percent of the probe packets from each of the three rounds (with a larger proportion in the first round) were re-scored to assess inter-scorer agreement. We computed an estimate of agreement by counting the number of problems considered agreements (i.e., scored correctly) and the number in which there was a disagreement in scoring (i.e., scoring errors) and dividing the number of agreements by the sum of agreements and disagreements. We computed the scoring accuracy by measure type for each of the selected students and then averaged across all students to obtain an overall estimate of inter-scorer agreement. The scoring accuracy results are presented in Table 3.

Data Analysis

Data analyses were conducted using number correct scores for the Basic Facts, MBSP-ConApp, Complex Quantity Discrimination, and Missing Number measures. Digits correct scores were used for the MBSP-Comp measure. For the Estimation measure, the final score was determined by subtracting one-half the number of incorrect responses from the total number of correct responses. The Basic Facts, Estimation, Complex Quantity Discrimination, and Missing Number probes were hand scored using answer keys. The MBSP measures were entered into the software program that is provided by the publisher and the resulting scores were entered into our database. For the criterion measures, teacher ratings were standardized by classroom and the resulting *z*-scores were used in the analyses. For the ITBS and NALT, we conducted analyses using national percentile rank scores (ITBS) and RIT scores (NALT).

Table 3

Mean Agreement, Range and Number of Probes Examined for Inter-Scorer Agreement

	MBSP Computation			MBSP Concepts & Applications		
	Mean Agreement	Range	# Probes Rescored	Mean Agreement	Range	# Probes Rescored
Fall	0.974	0.85-1.00	136	0.99	0.91-1.00	150
Winter	0.990	0.83-1.00	32	0.990	0.92-1.00	34
Spring	0.974	0.77-1.00	44	0.990	0.67-1.00	48
	Basic Facts			Estimation		
	Mean Agreement	Range	# Probes Rescored	Mean Agreement	Range	# Probes Rescored
Fall	0.997	0.93-1.00	230	0.991	0.38-1.00	282
Winter	0.999	0.97-1.00	32	0.998	0.90-1.00	66
Spring	0.997	0.96-1.00	72	0.991	0.81-1.00	90
	Complex Quantity Discrimination			Missing Number		
	Mean Agreement	Range	# Probes Rescored	Mean Agreement	Range	# Probes Rescored
Fall	0.998	0.92-1.00	336	0.994	0.67-1.00	336
Winter	0.999	0.96-1.00	78	0.997	0.94-1.00	78
Spring	0.998	0.92-1.00	105	0.994	0.80-1.00	105

We computed Pearson product moment correlation coefficients in our analyses of reliability and criterion validity. We estimated growth on the measures by determining Ordinary Least Squares regression slopes across the fall, winter, and spring data points and dividing the obtained slope values by the number of weeks between each data collection period to estimate weekly slope values.

Results

This section begins with descriptive statistics for all study measures. We next move to analyses specific to each of the research questions. Means and standard deviations for each of the individually administered middle school mathematics measures by grade level are presented in Table 4. Tests of skewness and kurtosis were conducted for all study variables and distributions met the assumptions for use of Pearson product moment correlations in our subsequent analyses of reliability and validity.

Table 4

Descriptive Statistics for Middle School Mathematics Measures

Grade 6							
Measure	Period	Measure	n	Min	Max	M	SD
MBSP Computation	Fall	Form A	68	16	57	35.51	9.03
		Form B	68	10	52	32.91	10.50
	Winter	Form A	70	13	73	43.06	13.88
		Form B	70	7	66	39.13	13.94
	Spring	Form A	70	15	77	45.16	13.57
		Form B	70	3	71	41.44	16.48
MBSP Concepts & Applications	Fall	Form A	83	0	28	12.94	5.90
		Form B	83	2	28	12.89	5.98
	Winter	Form A	84	4	33	17.76	7.19
		Form B	84	3	31	17.25	7.12
	Spring	Form A	82	5	35	19. 83	6.57
		Form B	82	6	35	20.27	6.98
Basic Facts	Fall	Form A	121	8	53	24.95	7.51
		Form B	121	8	45	25.25	7.83
	Winter	Form A	121	13	62	28.77	9.33
		Form B	121	9	63	28.40	9.63
	Spring	Form A	119	16	66	30.92	9.47
		Form B	119	12	62	31.18	10.31

Table 4 (continued)

Descriptive Statistics for Middle School Mathematics Measures

Measure	Period	Measure	n	Min	Max	M	SD
Estimation	Fall	Form A	111	0	28	11.85	5.48
		Form B	110	0	34	12.54	5.72
	Winter	Form A	111	2	35	17.73	6.30
		Form B	111	3	35	18.50	6.81
	Spring	Form A	114	3	38	20.58	7.74
		Form B	114	1	39	20.24	8.11
Complex Quantity Discrimination	Fall	Form A	81	6	28	14.95	4.73
		Form B	81	7	28	17.37	4.89
		Form C	81	6	3	18.31	5.30
	Winter	Form A	78	3	31	17.51	6.23
		Form B	78	6	35	21.12	6.30
		Form C	78	5	35	22.92	6.49
	Spring	Form A	81	7	42	21.01	6.78
		Form B	81	12	39	24.27	6.33
		Form C	81	13	42	25.86	6.41
Missing Number	Fall	Form A	81	1	16	7.85	3.53
		Form B	81	4	19	10.33	3.06
		Form C	81	1	16	8.79	3.41
	Winter	Form A	78	3	21	10.37	3.79
		Form B	78	7	25	11.95	3.09
		Form C	78	2	18	10.54	3.70
	Spring	Form A	81	3	21	11.84	4.14
		Form B	81	3	23	13.68	3.62
		Form C	81	3	23	12.02	3.95

Table 4 (continued)

Descriptive Statistics for Middle School Mathematics Measures

Grade 7							
Measure	Period	Measure	n	Min	Max	M	SD
MBSP Computation	Fall	Form A	73	23	58	38.47	8.09
		Form B	73	11	56	36.62	8.68
	Winter	Form A	71	26	66	45.62	9.31
		Form B	71	18	73	41.63	10.97
	Spring	Form A	74	15	72	45.93	12.15
		Form B	74	18	73	42.09	11.64
MBSP Concepts & Applications	Fall	Form A	77	4	37	18.22	8.18
		Form B	77	2	37	18.71	7.92
	Winter	Form A	73	8	39	23.19	8.96
		Form B	73	3	40	23.86	9.01
	Spring	Form A	72	7	39	25.10	8.17
		Form B	70	6	38	26.16	8.16
Basic Facts	Fall	Form A	128	9	56	30.57	9.90
		Form B	128	7	54	31.38	9.11
	Winter	Form A	123	9	71	35.41	11.47
		Form B	123	13	64	33.93	11.16
	Spring	Form A	120	14	73	36.85	11.31
		Form B	120	13	70	34.93	11.04
Estimation	Fall	Form A	126	0	32.5	12.06	5.22
		Form B	126	1	34	12.90	5.38
	Winter	Form A	124	7	34	17.49	5.20
		Form B	124	3	34	18.95	6.14
	Spring	Form A	119	5	36	19.57	5.95
		Form B	118	4	34	20.47	6.02

Table 4 (continued)

Descriptive Statistics for Middle School Mathematics Measures

Complex Quantity Discrimination	Fall	Form A	103	4	33	16.83	6.02
		Form B	103	3	33	19.64	5.97
		Form C	103	2	38	21.28	6.72
	Winter	Form A	102	5	44	23.03	7.14
		Form B	102	9	42	25.73	7.39
		Form C	102	9	44	26.91	7.17
	Spring	Form A	94	7	44	25.94	7.66
		Form B	94	10	44	28.45	7.96
		Form C	94	10	44	29.67	4.82
Missing Number	Fall	Form A	103	4	21	9.37	3.88
		Form B	103	0	26	11.18	4.23
		Form C	103	3	21	10.45	3.76
	Winter	Form A	103	5	26	11.74	4.81
		Form B	103	0	25	13.00	4.18
		Form C	103	0	25	12.16	4.76
	Spring	Form A	94	4	28	13.09	5.05
		Form B	94	6	27	14.43	4.47
		Form C	94	3	25	13.24	4.36

Table 4 (continued)

Descriptive Statistics for Middle School Mathematics Measures

Grade 8							
Measure	Period	Measure	n	Min	Max	M	SD
Estimation	Fall	Form A	57	1.5	27.5	15.36	6.17
		Form B	56	2	31	15.41	6.27
	Winter	Form A	53	10	32	20.43	5.45
		Form B	53	7	34	22.17	6.13
	Spring	Form A	55	8	39	22.91	7.06
		Form B	55	7	38	23.65	6.94
Complex Quantity Discrimination	Fall	Form A	57	7	31	16.81	5.14
		Form B	57	12	34	20.86	4.93
		Form C	57	10	38	23.49	5.87
	Winter	Form A	53	11	42	22.38	7.25
		Form B	53	14	43	26.42	7.27
		Form C	53	12	43	27.55	7.92
	Spring	Form A	55	8	42	24.38	8.00
		Form B	55	11	42	28.22	7.97
		Form C	55	12	44	29.42	7.71
Missing Number	Fall	Form A	57	3	27	9.58	4.67
		Form B	57	3	25	11.95	3.74
		Form C	57	3	28	10.21	4.78
	Winter	Form A	53	4	25	11.08	4.43
		Form B	53	1	21	12.36	4.10
		Form C	53	1	24	11.23	4.48
	Spring	Form A	55	1	29	12.22	5.26
		Form B	55	1	27	13.04	4.52
		Form C	55	0	26	11.87	5.07

In reviewing the data in Table 4, we considered the nature of the distributions produced by each measure. We were particularly interested in any floor or ceiling effects and the size of the standard deviations. Some Grade 6 students obtained scores of zero during the fall data collection period; this occurred for the MBSP-ConApp and Estimation measures. Some Grade 7 students obtained scores of zero during the fall and winter data collections periods; this occurred for the Estimation and Missing Number measures. Some Grade 8 students obtained scores of zero during the spring data collection period; this occurred for the Missing Number measure. As we examined the data more closely, we found that only one student at each grade level obtained a zero score among the Grades 6 and 7 participants. Two Grade 8 students obtained zero scores during the spring data collection period for the Missing Number measure.

We examined the standard deviations produced for each measure and found that the MBSP-Comp in Grades 6 and 7 and the Basic Facts in Grade 7 produced a much wider distribution of scores than did the other measures. In addition, students' mean scores on both measures and at the Grades 6 and 7 increased from one administration to the next.

Descriptive statistics for the criterion measures are presented in Table 5. ITBS scores are reported in percentile ranks, NALT scores as RIT scores, and Teacher Ratings as *z* scores, standardized by teacher.

Intercorrelations between the middle school mathematics measures are presented in Tables 6, 7, and 8. Positive relations were observed between all measures in Grade 6. The greatest intercorrelations were obtained between the MBSP-ConApp and the Basic Facts measures, while the weakest relations were observed between the Estimation and Complex Quantity Discrimination tasks. In Grade 7, a different pattern of results emerged, with the strongest intercorrelations produced by the Missing Number measure (with Basic Facts and

Estimation) and the weakest intercorrelations occurring between the MBSP-Comp and Estimation measures. Positive relations were observed between all measures in Grade 8. Intercorrelations were the strongest between the Complex Quantity Discrimination and the Missing Number measures.

Table 5

Descriptive Statistics for Criterion Variables

	Measure	n	Min	Max	M	SD
		Grade 6				
ITBS	Probability & Data	240	21	99	75.46	21.02
	Concepts & Estimation	239	13	99	76.38	19.99
	Total Math	236	21	99	76.77	19.86
NALT	Math, Fall	230	184	251	223.10	10.98
	Math, Spring	237	199	265	233.03	12.07
	Reading, Fall	234	189	243	219.26	9.82
	Reading, Spring	236	180	247	223.39	9.52
Teacher	Rating	235	-3.07	2.0	0.00	1.00
		Grade 7				
ITBS	Probability & Data	262	5	99	71.13	22.59
	Concepts & Estimation	262	7	99	72.26	21.44
	Total Math	262	11	99	72.27	21.62
NALT	Math, Fall	259	174	259	229.49	12.14
	Math, Spring	248	189	265	236.12	12.79
	Reading, Fall	263	168	245	221.63	11.32
	Reading, Spring	254	172	247	223.74	11.07
Teacher	Rating	256	-3.01	1.880	0.00	1.00
		Grade 8				
ITBS	Probability & Data	57	7	99	53.16	22.46
	Concepts & Estimation	57	18	94	57.42	21.44
	Computation	57	2	98	46.40	27.01
	Total Math	57	20	97	55.26	20.79
Teacher Rating		57	-2.13	1.42	0.00	1.00

Table 6

Intercorrelation Matrix for Middle School Math Measures at Grade 6

Measure	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
Grade 6																		
MBSP Computation																		
1 Fall	--																	
2 Winter	.84**	--																
3 Spring	.77**	.79**	--															
MBSP Concepts & Applications																		
4 Fall				--														
5 Winter				.86**	--													
6 Spring				.84**	.87**	--												
Basic Facts																		
7 Fall				.64**	.70**	.60**	--											
8 Winter				.63**	.67**	.57**	.90**	--										
9 Spring				.65**	.66**	.60**	.92**	.93**	--									
Estimation																		
10 Fall	.56**	.50**	.51**							--								
11 Winter	.56**	.51**	.53**							.74**	--							
12 Spring	.60**	.54**	.59**							.75**	.81**	--						
Complex Quantity Discrimination																		
13 Fall							.48**	.47**	.44**	.29	.16	.22	--					
14 Winter							.61**	.64**	.54**	.27	.17	.29	.80**	--				
15 Spring							.63**	.63**	.61**	.24	.06	.38*	.67**	.81**	--			
Missing Number																		
16 Fall							.53**	.56**	.59**	.46**	.18	.46**	.64**	.65**	.52**	--		
17 Winter							.49**	.63**	.61**	.39*	.26	.40*	.53**	.63**	.46**	.79**	--	
18 Spring							.58**	.62**	.62**	.45**	.19	.51**	.54**	.60**	.53**	.80**	.82**	--

Note. * = $p < .05$, ** = $p < .01$. Numbers at the top of the table correspond to variables listed in far left column.

Table 7

Intercorrelation Matrix for Middle School Math Measures at Grade 7

Measure	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
Grade 7																		
MBSP Computation																		
1 Fall	--																	
2 Winter	.51**	--																
3 Spring	.42**	.78**	--															
MBSP Concepts & Applications																		
4 Fall				--														
5 Winter				.94**	--													
6 Spring				.90**	.94**	--												
Basic Facts																		
7 Fall				.64**	.64**	.66**	--											
8 Winter				.66**	.66**	.67**	.93**	--										
9 Spring				.67**	.64**	.65**	.88**	.94**	--									
Estimation																		
10 Fall	.22	.17	.10							--								
11 Winter	.27*	.26*	.24*							.72**	--							
12 Spring	.17	.01	.12							.61**	.73**	--						
Complex Quantity Discrimination																		
13 Fall							.57**	.52**	.45**	.55**	.49**	.31*	--					
14 Winter							.54**	.48**	.56**	.50**	.67**	.60**	.81**	--				
15 Spring							.61**	.63**	.63**	.45**	.60**	.64**	.75**	.94**	--			
Missing Number																		
16 Fall							.72**	.68**	.66**	.73**	.70**	.69**	.53**	.51**	.50**	--		
17 Winter							.76**	.76**	.80**	.62**	.67**	.68**	.54**	.59**	.60**	.86**	--	
18 Spring							.71**	.77**	.75**	.65**	.66**	.73**	.49**	.61**	.66**	.82**	.90**	--

Note. * = p < .05, ** = p < .01. Numbers at the top of the table correspond to variables listed in far left column.

Table 8

Intercorrelations between Middle School Mathematics Measures for Grade 8

Measure	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
Grade 8																		
MBSP Computation																		
1 Fall	--																	
2 Winter		--																
3 Spring			--															
MBSP Concepts & Applications																		
4 Fall				--														
5 Winter					--													
6 Spring						--												
Basic Facts																		
7 Fall							--											
8 Winter								--										
9 Spring									--									
Estimation																		
10 Fall										--								
11 Winter										.65**	--							
12 Spring										.71**	.82**	--						
Complex Quantity Discrimination																		
13 Fall										.42**	.32*	.42**	--					
14 Winter										.49**	.43*	.55**	.83**	--				
15 Spring										.46**	.47**	.60**	.81**	.90**	--			
Missing Number																		
16 Fall										.48**	.15	.34*	.52**	.66**	.63**	--		
17 Winter										.52**	.47**	.59**	.61**	.70**	.74**	.81**	--	
18 Spring										.57**	.35*	.54**	.63**	.72**	.75**	.85**	.80**	--

Note. * = $p < .05$, ** = $p < .01$. Numbers at the top of the table correspond to variables listed in far left column.

Intercorrelations between the criterion variables are presented in Table 9. We observed positive relations between all of the criterion measures.

Table 9

Intercorrelations Between the Criterion Measures

	Measure	1	2	3	4	5	6	7
	Grade 6							
1	Teacher Rating, Fall	--						
2	NALT Math, Fall	.75	--					
3	NALT Reading, Fall	.64	.68	--				
4	NALT Math, Spring	.78	.82	.64	--			
5	NALT Reading, Spring	.56	.59	.73	.60	--		
ITBS (Winter)								
6	Total Math	.72	.82	.61	.78	.55	--	
7	Probability & Data	.66	.75	.58	.70	.54	.94	--
8	Concept Estimation	.71	.76	.54	.78	.50	.90	.73
	Grade 7							
1	Teacher Rating, Fall	--						
2	NALT Math, Fall	.66	--					
3	NALT Reading, Fall	.54	.72	--				
4	NALT Math, Spring	.66	.86	.68	--			
5	NALT Reading, Spring	.55	.72	.82	.73	--		
ITBS (Winter)								
6	Total Math	.59	.81	.67	.82	.70	--	
7	Probability & Data	.53	.73	.63	.75	.65	.95	--
8	Concept Estimation	.58	.80	.62	.80	.66	.91	.75
	Grade 8							
1	Teacher Rating, Fall	--						
ITBS (Winter)								
2	Total Math	.55	--					
3	Probability & Data	.42	.94	--				
4	Concept Estimation	.62	.88	.66	--			
5	Computation	.61	.66	.52	.73	--		

Note: All correlations significant, $p < .01$

Research Question 1: What is the reliability of the middle school mathematic measures?

Alternate-form reliability. We examined alternate-form reliability by computing Pearson product moment correlation coefficients between the scores students obtained on the two forms of the MBSP-Comp, MBSP-ConApp, Basic Facts, and Estimation measures. For the Complex

Quantity Discrimination Measure and the Missing Number measure (for which students completed three forms of each measure), we computed the correlations between each pair of forms and reported the mean of these values. Table 10 displays the results of the alternate-form reliability analyses. Only one of the measures, Basic Facts, met the conventional benchmark of $r = .80$ consistently across administration periods and grade levels. The MBSP-ConApp and the Complex Quantity Discrimination missed the criterion for Grade 6 students in the fall (r = .76 and .77, respectively), but met it for all other data points. The Estimation measure was more reliable for students in Grades 6 and 8 than for students in Grade 7. The Missing Number measure consistently produced the lowest estimated of alternate-form reliability across all three grade levels. The MBSP-Comp produced acceptable levels of reliability for Grade 6 students in the winter and spring, but was far below desirable levels for Grade 7 students in the fall and in the winter.

Test retest reliability. Test retest reliability correlation coefficients are reported in Table 11. We first considered the degree to which the measures were reliable when a student's score from a single measure was correlated to his/her score from that same measure administered one week later. These data appear under the columns labeled Form A, B, and C in Table 11. Two of the measures, MBSP-ConApp and Basic Facts, produced coefficients that exceeded the conventional $r = .80$ level across all forms and all grade levels. For the remaining measures, the results varied by grade level. The MBSP-Comp produced particularly low levels of test-retest reliability for Grade 7 students ($r = .48$ and .34). For the Estimation measure, scores for Grade 6 students met the $r = .80$ criterion, but fell below for those students in Grades 7 and 8. Both the Complex Quantity Discrimination measure and the Missing Number measure produced stronger coefficients at Grades 7 and 8 than at Grade 6.

Table 10

Alternate-Form Reliability of Middle School Progress Monitoring Measures

	Grade 6		Grade 7		Grade 8	
	N	*r*	N	*r*	N	*r*
MBSP Comp						
Fall	68	.64	73	.43		
Winter	70	.81	71	.57		
Spring	70	.84	74	.78		
MBSP ConApp						
Fall	83	.76	77	.86		
Winter	84	.80	73	.88		
Spring	82	.82	70	.81		
Basic Facts						
Fall	121	.80	128	.85		
Winter	121	.91	123	.91		
Spring	119	.89	120	.90		
Estimation						
Fall	110	.81	126	.78	56	.78
Winter	111	.83	124	.70	53	.90
Spring	114	.84	118	.72	55	.80
CQD						
Fall	81	.77	103	.82	57	.80
Winter	78	.87	102	.87	55	.88
Spring	81	.86	94	.88	55	.89
Missing Number						
Fall	81	.65	103	.66	57	.68
Winter	78	.69	103	.74	53	.58
Spring	81	.72	94	.79	55	.80

Note: All correlations significant, $p < .01$

Table 11

Test Retest Reliability of Single form and Aggregated Scores on Middle School Math Measures

Measure	Grade	Form A	Form B	Form C	Avg. of 2 AB	Avg. of 2 BC	Avg. of 2 AC	Avg. of 3	Median of 3
MBSP-Comp	6	.73	.70	--	.85				
	n	*65*	*66*		*66*				
	7	.48	.34	--	.55				
	n	*72*	*72*		*72*				
MBSP-ConApp	6	.88	.83	--	.92				
	n	*82*	*82*		*82*				
	7	.93	.90	--	.95				
	n	*75*	*75*		*75*				
Basic Facts	6	.84	.82	--	.90				
	n	*120*	*120*		*120*				
	7	.88	.88	--	.92				
	n	*126*	*126*		*126*				
Estimation	6	.80	.81	--	.86				
	n	*103*	*103*		*103*				
	7	.69	.67	--	.75				
	n	*118*	*118*		*118*				
	8	.70	.67	--	.80				
	n	*55*	*54*		*55*				
Complex Quantity Discrim.	6	.77	.65	.67	.82	.71	.83	.82	.75
	n	*77*	*77*	*77*	*77*	*77*	*77*	*77*	*77*
	7	.71	.80	.75	.82	.82	.81	.84	.84
	n	*96*	*96*	*96*	*96*	*96*	*96*	*96*	*96*
	8	.76	.85	.84	.86	.90	.87	.90	.87
	n	*55*	*55*	*55*	*55*	*55*	*55*	*55*	*55*
Missing Number	6	.69	.72	.74	.82	.82	.78	.84	.78
	n	*76*	*77*	*77*	*77*	*77*	*77*	*77*	*77*
	7	.82	.69	.77	.83	.83	.86	.88	.85
	n	*96*	*96*	*96*	*97*	*96*	*97*	*97*	*97*
	8	.85	.79	.85	.88	.90	.90	.92	.85
	n	*55*	*55*	*55*	*55*	*55*	*55*	*55*	*55*

Note: All correlations significant, $p < .01$

We next considered the reliability of across testing sessions when scores from two alternate forms were first averaged before the correlations were computed. In all cases, this aggregation increased the reliability of the scores. With the exception of the MBSP-Comp at Grade 7, all of the coefficients exceeded the *r* = .80 level when using this method.

For the two measures for which students completed three forms (Complex Quantity Discrimination and Missing Number) we could also consider the effects of aggregating across the mean or the median of three forms when evaluating test-retest reliability. These results are presented in the last two columns of Table 11. When three forms were aggregated, we obtained small or negligible improvements in the reliability of the measures over the results we obtained when using the mean of two forms.

Research Question 2: What is the criterion validity of the middle school mathematics measures?

The results of the criterion validity analyses are reported separately by data collection period. In Tables 12, 13, and 14, we report the concurrent validity coefficients for each grade level for the fall, winter, and spring data collection periods, respectively. In Table 15, we report the predictive validity coefficients for each grade level. Criterion validity coefficients were calculated using the mean of two probes. The criterion measures used in the study included teachers' ratings of students' overall math proficiency, standardized test scores (ITBS), and a comprehensive assessment associated with the district mathematics curriculum (NALT). In general, most of the concurrent validity coefficients were in the moderate to strong range, particularly with respect to the strength of coefficients typically obtained for mathematics measures (cf., Foegen, Jiban, & Deno, 2007). For the fall administration (Table 12), validity coefficients for MBSP-ConApp were consistently higher (*r* = .73 to .86) than those obtained for the other five measures at Grade 6 and 7 levels. The weakest criterion validity relations were identified between the MBSP-Comp at Grade 7 and Estimation at Grades 7 and 8.

Table 12

Concurrent Validity of Middle School Measures for Fall

Measure	Grade	Teacher Ratings		NALT Math		NALT Reading	
		N	*r*	N	*r*	N	*r*
MBSP-Comp	6	67	.65	65	.64	67	.59
	7	73	.35	71	.38	73	.40
MBSP ConApp	6	82	.75	79	.76	81	.58
	7	77	.73	76	.86	77	.68
Basic Facts	6	119	.56	116	.50	118	.39
	7	127	.54	126	.60	128	.33
Estimation	6	108	.60	106	.57	107	.45
	7	120	.26	124	.45	126	.32
	8	57	.40				
Complex	6	78	.57	78	.52	77	.37
Quantity	7	96	.42	102	.57	103	.37
Discrimination	8	57	.51				
Missing Number	6	78	.56	78	.47	77	.27*
	7	96	.42	102	.67	103	.39
	8	57	.50				

Note: All correlations significant unless marked, $p < .01$; * = $p < .05$

For the winter administration (Table 13), concurrent criterion validity coefficients for the MBSP-ConApp were again consistently higher than those obtained for the other measures at both Grades 6 and 7. In general, the progress monitoring measures tended to be more closely related to the ITBS Concepts and Estimation subtest scores and to the ITBS Total scores than to the ITBS Problems and Data subtest scores. A pattern similar to the fall data emerged for

Table 13

Concurrent Validity of Middle School Measures for Winter

Measure	Grade	ITBS Problems & Data		ITBS Concepts & Estimation		ITBS Math Total	
		N	*r*	N	*r*	N	*r*
MBSP-Comp	6	69	.54	69	.60	69	.59
	7	71	.33	71	.34	71	.38
MBSP ConApp	6	84	.60	84	.72	84	.71
	7	73	.80	73	.87	73	.87
Basic Facts	6	121	.38	121	.52	119	.49
	7	123	.52	123	.55	123	.55
Estimation	6	109	.50	109	.51	108	.53
	7	124	.47	124	.46	124	.51
	8	53	.53	53	.52	53	.43
Complex	6	77	.50	77	.52	74	.53
Quantity	7	102	.57	102	.57	102	.60
Discrimination	8	53	.52	53	.41	53	.55
Missing Number	6	77	.45	77	.45	74	.46
	7	103	.51	103	.53	103	.54
	8	53	.41	53	.31	53	.47

Note: All correlations significant, $p < .01$

the MBSP-Comp, with the Grade 7 coefficients substantially lower than those obtained for Grade 6. The weakest relations among the winter data were with the Missing Number measure across all three grade levels and for the MBSP-Comp at Grade 7.

For the spring administration (Table 14), progress monitoring scores for students in Grades 6 and 7 were correlated with their scores from the spring administration of the NALT. The strongest relations for students in Grade 7 were with the MBSP-ConApp, Missing Number, and Complex Quantity Discrimination measures. For Grade 6 students, the measures with the strongest criterion validity in the spring were the MBSP-Comp, Estimation, and MBSP-ConApp.

The same pattern of differential results for the MBSP-Comp occurred in the spring data, with the relation for Grade 6 students being the strongest from among the six options, while the correlation coefficient for Grade 7 was both negligible (.18) and nonsignificant.

Table 14

Concurrent Validity of Middle School Measures for Spring

Measure	Grade	NALT Math		NALT Reading	
		N	*r*	N	*r*
MBSP-Comp	6	69	.64**	67	.60**
	7	69	.18	72	.29*
MBSP ConApp	6	69	.57**	67	.50**
	7	71	.87**	72	.68**
Basic Facts	6	116	.49**	117	.29**
	7	117	.58**	119	.33**
Estimation	6	112	.59**	110	.52**
	7	113	.29**	117	.34**
Complex Quantity	6	79	.31**	79	.03**
Discrimination	7	91	.61**	93	.42**
Missing Number	6	79	.46**	79	.21**
	7	91	.63**	93	.41**

Note: * = $p < .05$, ** = $p < .01$

Predictive validity coefficients for Grade 6 and 7 students were computed by correlating students' fall scores on the progress monitoring measures with their spring scores from the NALT (see Table 15). The correlation coefficients for the MBSP-ConApp were the strongest among the six measures for students in both grade levels. For 6th grade students, the next strongest relations were with the MBSP-Comp, followed by Basic Facts, Estimation, and Complex Quantity Discrimination. For Grade 7 students, the next strongest measure was Missing Number (which had the weakest relations for Grade 6 students), followed by Basic Facts, Estimation, and Complex Quantity Discrimination.

Table 15

Predictive Validity of Fall Middle School Math Measures to Spring Criterion Variables

Measure	Grade	NALT Math		NALT Reading	
		N	*r*	N	*r*
MBSP-Comp	6	67	.64**	65	.62**
	7	69	.25*	71	.23
MBSP ConApp	6	83	.76**	82	.44**
	7	76	.87**	77	.65**
Basic Facts	6	118	.55**	119	.24**
	7	122	.59**	124	.36**
Estimation	6	109	.55**	107	.52**
	7	119	.34**	122	.41**
Complex Quantity	6	79	.53**	79	.25*
Discrimination	7	95	.58**	97	.44**
Missing Number	6	79	.48**	79	.22
	7	95	.60**	97	.39**

Note: * = $p < .05$, ** = $p < .01$

Across all of the criterion validity data, the strength of the MBSP-ConApp is quite remarkable. It demonstrated the strongest relations with a variety of criterion measures in both concurrent and predictive situations. The differential pattern of results for the MBSP-Comp is also noteworthy, given the relative strength of the relations between this measure and criterion measures for students in Grade 6, and the relative absence of any meaningful relations for students in Grade 7.

The correlations with NALT Reading also offer some interesting patterns in the criterion validity data. Three measures (MBSP-Comp, MBSP-ConApp, and Estimation) produced correlations with the NALT Reading that were either substantial in comparison to correlations obtained by other mathematics progress monitoring measures with other criterion variables, or in some instances, stronger than the relations for the same measure with the NALT Math. This finding is less surprising for the MBSP-ConApp and the Estimation measures, which both require reading to interpret and respond to the items. We did not anticipate this result for the MBSP-Comp measure.

Research Question 3: How much academic growth do students display on each of the measures over the course of one academic school year?

Our final research question examined the extent to which students' scores on the middle school progress monitoring measures changed over time. Our first set of analyses used effect sizes for each of the measures. We computed a difference score (spring average minus fall average) for each student and expressed the mean of these scores within a grade level and by teacher in standard deviation units. In Table 16 we report the average effect sizes for each of the six progress monitoring measures. The largest effect sizes (near or exceeding two standard deviations) were obtained for MBSP-ConApp measure. The next largest effect sizes (near or exceeding 1.5 standard deviations) were obtained for the Complex Quantity Discrimination measure at all three grade levels, and the Basic Facts and Missing Number measures at Grade 6. The smallest effect sizes for Grades 6 and 7 were obtained for the MBSP-Comp measure, while the Estimation and Missing Number measures had considerably smaller effect sizes for students in Grade 8.

Table 16

Standardized Within-Grade Growth on Middle School Math Measures

	Grade 6		Grade 7		Grade 8	
Measure	n	ES	n	ES	n	ES
MBSP-Comp	66	0.88	71	0.70		
MBSP-ConApp	79	2.13	71	1.93		
Basic Facts	116	1.45	118	0.93		
Estimation	108	1.19	114	1.12	55	0.80
Complex Quantity Discrimination	78	1.46	90	1.68	55	1.49
Missing Number	78	1.62	90	1.25	55	0.79

Another method for evaluating a measure's sensitivity to changes in student performance over time is the slope of a line calculated using the data points. The data in Table 17 represent slope values calculated by computing ordinary least squares (OLS) regression through each student's data points. Students who had data for at least two of the three points were included in these analyses. Slope values were adjusted to reflect expected weekly rates of growth and are reported in the scoring unit for each measure, rather than in standardized form. In other words, the slope of .41 for the MBSP-Comp in Grade 6 can be interpreted as representing an average increase of .41 digits correct per week (the scoring metric for MBSP-Comp) among Grade 6 students. The SEE column represents the mean Standard Error of the Estimate, or the amount of variability around the regression line. An SEE that is substantially larger than the slope indicates a considerable degree of "bounce" in the data points around the regression line, which would likely make interpreting the data more difficult. This factor is constrained in the current data set because of the small number of points used to generate the regression line.

Table 17

Weekly Slope Values by Grade Level

	Grade 6			Grade 7			Grade 8		
Measure	N	Slope	SEE	N	Slope	SEE	N	Slope	SEE
MBSP-Comp	72	.41	.26	91	.36	.21			
MBSP-ConApp	91	.42	.27	76	.31	.11			
Basic Facts	122	.27	.13	126	.22	.12			
Estimation	117	.28	.13	126	.22	.12	56	.20	.16
Complex Quantity Discrimination	83	.31	.11	103	.38	.12	56	.31	.13
Missing Number	83	.16	.06	103	.13	.06	56	.09	.07

For students in Grades 6 and 7, slope values were greatest for the MBSP-Comp, MBSP-ConApp and Complex Quantity Discrimination measures. In Grade 8, the Complex Quantity Discrimination measure produced the largest mean slope value. In each case, these values exceeded 0.30 points per week of improvement. The Missing Number measure consistently produced the smallest weekly slope values, with rates at or below 0.16 points per week at all three grade levels.

Discussion and Future Research

The purpose of this study was to examine the technical adequacy of six potential measures for monitoring student progress in mathematics at the middle school level. We examined two measures used in previous middle school research (Basic Facts, Estimation), two measures with established technical adequacy for progress monitoring at the elementary grades (MBSP-Comp, MBSP-ConApp) and two new measures (Complex Quantity Discrimination, Missing Number). Five hundred sixty three students in Grades 6, 7, and 8 completed two or more types of probes at three points during the 2005-06 school year. Criterion measures included teacher ratings gathered in the fall and spring, final course grades (Grade 8 only) and scores on standardized achievement tests administered in the participating districts. Our analyses were guided by three research questions addressing the reliability, criterion validity, and sensitivity to growth of the potential measures.

With respect to reliability, we found that most of the measures met expected benchmarks for adequate alternate-form reliability, though scores from the fall administration tended to produce lower coefficients. The Missing Number measure consistently fell below expected benchmarks for most grade levels at most time periods, as did the MBSP-Comp for students in Grade 7. When we examined test-retest reliability and considered the benefits of combining scores from multiple forms, we found that two of the measures (MBSP-ConApp and Basic Facts) consistently produced reliable scores, even when the correlations were based on only a single form. Among the remaining measures, using the average of two forms produced correlation coefficients at or near the conventional benchmark of $r = .80$. One exception was the MBSP-Comp at Grade 7, for which we obtained an extremely low correlation coefficient even when averaging across two forms. For two of the measures we examined the use of the mean and median of three forms and found no benefits beyond those obtained for the mean of two forms.

We examined both concurrent and predictive validity for the measures. The strongest concurrent criterion validity coefficients were obtained for the MBSP-ConApp in Grades 6 and 7 and the Complex Quantity Discrimination measure for students in Grade 8 (who did not complete either of the MBSP measures). Consistently weak relations were obtained with the MBSP-Comp for students in Grade 7. Predictive validity was examined with students in Grades 6 and 7; these results paralleled the concurrent validity results. For students in Grade 6, strong relations were obtained for the MBSP-ConApp and MBSP-Comp measures. For students in Grade 7, relations were strongest for the MBSP-ConApp measures (although the coefficients for three other measures were near $r = .60$: Missing Number, Basic Facts, and Complex Quantity Discrimination).

The final research question addressed the issue of growth over time on the measures. We examined effect sizes based on fall to spring difference scores as well as linear regression slopes based on the three data points. Effect sizes exceeding 1.4 were obtained for the MBSP-ConApp measure in Grades 6 and 7, the Complex Quantity Discrimination measure in all three grades, and the Basic Facts and Missing Number measures in Grade 6. We computed individual slope values and divided by the approximate number of weeks between data collection sessions to estimate weekly growth rates for each student. In Grade 6, the greatest levels of growth were obtained on the MBSP-ConApp and the MBSP-Comp (.42 and .41 points per week, respectively). In Grade 7, students demonstrated the most growth on the Complex Quantity Discrimination measure and the MBSP-Comp (.38 and .36 points per week, respectively). Finally, in Grade 8, students improved .31 points per week on the Complex Quantity Discrimination measure.

Based on these findings (which much be considered tentative, especially given the relatively small size of the Grade 8 sample), we would offer the following recommendations to practitioners. The most reliable estimates of student performance were obtained when we averaged across two forms of the measures. While a single form may prove to be sufficient in future research within a progress monitoring context, we would recommend using two forms and averaging students' scores in a benchmarking assessment program when the measures are only given two to four times a year. With regard to the measures that are most likely produce valid data and demonstrate sufficient growth over time, we would recommend the MBSP-ConApp for students in Grades 6 and 7 (as well as the MBSP-Comp in Grade 6) and the Complex Quantity Discrimination measure for students in Grade 8. Because the results of the validity analyses varied across criterion measures, we would also recommend that practitioners examine the criterion validity of the measures with respect to the important outcome measures in their schools, districts, and states.

Future research studies should replicate these findings with larger samples (particularly at Grade 8) and the use of the MBSP-ConApp and MBSP-Comp with Grade 8 students. It would also be valuable to extend the range of criterion measures and to explore relations between the measures and students' course grades, particularly in Grades 6 and 7. A natural next step in the research is to examine the use of the measures for monitoring student progress at more frequent intervals. The relations with measures of reading also pose an interesting area for future investigation.

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. *Journal of Special Education, 35*, 4-16.

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

Fuchs, L. S., Hamlett, C. L., & Fuchs, D. (1999). *Monitoring basic skills progress: Basic math concepts and applications.* Austin, TX: Pro-Ed.

Fuchs, L. S., Hamlett, C. L., & Fuchs, D. (1998). *Monitoring basic skills progress: Basic math computation* (2nd Ed.). Austin, TX: Pro-Ed.

Heartland Area Education Agency 11 (n.d.). *Heartland AEA 11agency-wide curriculum-based measurement norms: 2002-2003*. Johnston, IA: Author.

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

Helwig, R., & Tindal, G. (2002). Using general outcome measures in mathematics to measure adequate yearly progress as mandated by Title I. *Assessment for Effective Intervention, 28*(1), 9-18.

Hoover, H. D., Dunbar, S. B., & Frisbie, D. A. (2001). *Iowa Tests of Basic Skills*. Chicago: Riverside Publishing Company.

Northwest Evaluation Association. (2002). *Northwest Achievement Levels Test*. Portland, OR: Author.

Northwest Evaluation Association. (Sept. 2003). *Technical manual for use with Measures of Academic Progress and Achievement Level Tests*. Portland, OR: Author.

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

Appendix A

Middle School Mathematics Measures

Monitoring Basic Skills Progress-Computation

Sample Measure

Administration Directions

Monitoring Basic Skills Progress-Concepts and Applications

Sample Measure (First page only)

Administration Directions

Basic Facts

Sample Measure

Administration Directions

Estimation

Sample Measure (First page only)

Administration Directions

Complex Quantity Discrimination

Sample Measure

Administration Directions

Missing Number

Sample Measure

Administration Directions

Sheet #1 Computation 6

Password: ARM

Name: ______________________ Date: ______________

A $\frac{3}{5} - \frac{1}{3} =$	B 2.66×5.4	C $5\frac{3}{5} - 3\frac{4}{5} =$	D $15961 + 92307$	E $23281 - 16754$
F $2.591 + 7.6588$	G $65983 + 56937$	H $.13\overline{)884}$	I $122\overline{)8614}$	J $3 \times \frac{1}{2} =$
K 5952×246	L $7\frac{4}{7} + 1\frac{2}{3} =$	M $45\overline{)65}$	N $3\frac{1}{3} + 8\frac{2}{3} =$	O $3.4423 - 1.33$
P $\frac{2}{5} \times \frac{2}{5} =$	Q $81\overline{)9301}$	R 1.292×1.7	S $1.3\overline{).598}$	T $\frac{7}{9} + \frac{2}{3} =$
U 3596×168	V $7 \div \frac{1}{5} =$	W $\frac{3}{4} \div \frac{7}{9} =$	X $9\frac{3}{10} - 3\frac{3}{5} =$	Y $55867 - 32719$

Research Institute Progress Monitoring (RIPM)
2005-2006
Middle School Measures Study

MBSP Basic Computation
Standard Directions

Set timer to 6 minutes.

*"**The first type of probe you will be doing is a Computation probe.*** [Hold up booklet with Computation sample page showing.] ***This sample page shows some examples of the types of problems on the probe. You can see that problem A has been answered. The dot is a symbol for multiplication. On some of the probes, you will see the dot used to stand for multiplication and on others, you will see an x, or 'times' sign. In problem A, two sevenths multiplied by five eighths is five 28ths. Notice that the answer has been written in lowest terms. You are welcome to show your work as you solve these problems. Now try problem B. When you do division problems on this probe, please write the remainder as a decimal and round to two places. Please take a minute to solve problem B.*** [Pause as students work.] ***What is the answer to problem B?*** [Wait for students to volunteer 1.04, or one and four hundredths]. ***Yes, that's correct.***

Here's how you take this test. Start here (point to the upper left corner) ***and move across each row from left to right*** (demonstrate). ***When you come to a problem that's easy for you, do it right away. When you come to a problem that's hard, skip it. When you've looked through the whole test and finished the easy problems, then go back to the beginning and try the harder ones.***

We will score these tests by looking at each number, or digit, in the answer. That way, you can earn partial credit even if you answer isn't completely correct. As you work the problems, please write fractions in lowest terms (reduce/simplify them). When you complete a division problem, write the remainder as a decimal and round your answer to two places (hundredths). You can convert improper fractions to mixed numbers OR leave them as improper fractions. Either form of the correct answer will earn full credit.

Now we'll take a minute so you can practice doing a Computation probe. If you finish before I say 'Stop', please do NOT turn to the next page. Any questions?

Ready, begin. [Time for 1 minute—watch clock or watch]

Stop. Please put your 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 of two Computation probes. You will have *6* ***minutes to work on this probe****.* ***Remember, your job is to answer as many problems correctly as you can in 6 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. If you get to the STOP page, remember to stop there.***

****When I say begin, please turn past the sample/STOP page and start working. Remember to start with the easy problems and then go back to the problems that are more difficult. You will have 6 minutes. Do your best work.***

Ready, (Pause) ***begin*****.** [Start timer. Time for 6 minutes.] When the timer goes off, say, ***Stop. Please put your pencils down. If you haven't turned to the red STOP page, please do so. [Now we will do the second Computation probe.]*** [REPEAT FROM * FOR SECOND PROBE.]

Now we will shift gears to a different type of probe.

Name ______________________________ Date ________________ Test 1 Page 1

Column A

Applications 6

Column B

(1)

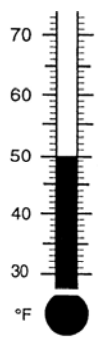
Write **P** if the number is a prime number and **C** if the number is a composite number.

____ 2 ____ 94

(2)

$7^2 =$ _____

(3)



last night's temperature

When Emily woke up, the temperature was 42° F. By how many degrees did the temperature fall?

________ ° F

(4) Which expression matches the phrase:

The difference between y and 19?

(A) $y - 19$

(B) $\frac{19}{y}$

__________ (C) $y + 19$

If $y = 25$, then the value of the expression is

____

A

(5)

Rename if necessary.

$$\begin{array}{r} 3 \text{ m} \quad 92 \text{ cm} \\ + \; 7 \text{ m} \quad 15 \text{ cm} \\ \hline \end{array}$$

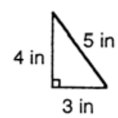
___m ___cm

(6)

15 girls wore pink dresses, 25 wore blue dresses, 7 wore purple dresses and 2 wore green dresses. Write the ratio of green dresses to purple dresses, using the word "to."

____

(7)



(A) acute

(B) obtuse

(C) right

What kind of triangle? ________

(8)

Express 7% as:

a decimal _____

a fraction with denominator of 100 _____

(9)

2:5 is the same as ___:15

Research Institute Progress Monitoring (RIPM)
2005-2006
Middle School Measures Study

MBSP Concepts and Applications
Standard Directions

Set timer to 7 minutes.

"***The second type of probe you will be doing is a Concepts and Applications probe.*** [Hold up booklet with Concepts and Applications sample page showing.] ***This sample page shows some examples of the types of problems on the probe. It has many different kinds of problems on it, including multiple choice problems, story problems, and problems where you need to read charts and graphs. Look at sample problem A. On this problem, you need to read the thermometer to answer the question. What do you think the correct answer is?*** [Pause and allow students to work out the problem; solicit a volunteer to answer.] ***Yes, that's correct. The temperature started at 4 degree and increased to 12 degrees. The number of degrees it increased is 8. Now try problem B.*** [Pause as students work.] ***What is the answer to problem B?*** [Wait for students to volunteer $142.58]. ***Yes, that's correct.***

Here's how you take this test. Start here (point to the upper left corner) ***and move down each column*** (demonstrate). ***When you come to a problem that's easy for you, do it right away. When you come to a problem that's hard, skip it. When you've looked through the whole test and finished the easy problems, then go back to the beginning and try the harder ones.***

We will score these tests by looking at each answer you give. As you work the problems, please write fractions in lowest terms (reduce/simplify them). When you do a problem that involves money, remember to use a dollar sign in your answer.

Now we'll take a minute so you can practice doing a Concepts and Applications probe. If you finish before I say 'Stop', please do NOT turn to the next page. Any questions?

Ready, begin. [Time for 1 minute—watch clock or watch]

Stop. Please put your 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 of two Concepts and Applications probes. You will have *7* ***minutes to work on this probe****.* ***Remember, your job is to answer as many problems correctly as you can in 7 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. If you get to the STOP page, remember to stop there.***

When I say begin, please turn past the sample/STOP page and start working. Remember to start with the easy problems and then go back to the problems that are more difficult. You will have 7 minutes. Do your best work.*

Ready, (Pause) ***begin*****.** [Start timer. Time for 7 minutes.] When the timer goes off, say, ***Stop. Please put your pencils down. If you haven't turned to the red STOP page, please do so. [Now we will do the second Concepts and Applications probe.]*** [REPEAT FROM * FOR SECOND PROBE.]

That is the end of the probes for today. THANK YOU for your hard work!

Facts Probe 1

1 - 1 =	8 x 1 =	5 x 5 =	1 x 7 =
4 x 7 =	5 x 7 =	4 + 6 =	9 x 5 =
3 + 0 =	$3\overline{)12}$ =	$2\overline{)14}$ =	$6\overline{)6}$ =
12 - 9 =	7 + 4 =	0 x 7 =	7 - 4 =
$5\overline{)10}$ =	$8\overline{)48}$ =	11 - 7 =	$4\overline{)12}$ =
8 - 2 =	9 + 6 =	6 + 6 =	1 x 2 =
8 + 7 =	0 x 0 =	11 - 2 =	8 - 5 =
6 - 2 =	7 + 0 =	3 + 3 =	17 - 9 =
10 - 4 =	9 x 9 =	$4\overline{)4}$ =	$1\overline{)5}$ =
$1\overline{)1}$ =	2 -2 =	5 + 9 =	7 x 8 =
$6\overline{)54}$ =	9 - 3 =	$4\overline{)32}$ =	16 - 7 =
4 + 5 =	14 - 9 =	7 + 6 =	2 x 6 =
8 + 8 =	13 - 6 =	2 x 4 =	$5\overline{)0}$ =
1 + 0 =	6 x 2 =	2 + 8 =	1 + 8 =
$9\overline{)63}$ =	$3\overline{)27}$ =	$3\overline{)15}$ =	$9\overline{)36}$ =
0 + 0 =	8 x 3 =	8 + 5 =	$7\overline{)42}$ =
13 - 8 =	$6\overline{)24}$ =	2 x 2 =	2 - 0 =
9 + 1 =	6 - 3 =	0 + 7 =	3 x 5 =
$8\overline{)8}$ =	4 x 9 =	9 - 7 =	$5\overline{)40}$ =
5 + 2 =	7 - 0 =	1 x 6 =	8 + 0 =

Research Institute Progress Monitoring (RIPM)
2005-2006
Middle School Measures Study

Basic Facts
Standard Directions

Set timer to 1 minute.

“***The first type of probe you will be doing is a Basic Facts probe.*** [Hold up booklet with Basic Facts sample page showing.] ***This sample page shows some examples of the types of problems on the probe. For each problem, you will write the answer. You can see that problem A has been answered. 13 minus 9 is 4 and that is the answer given. If you look at problem B, you will see 7 and 9 with a dot between them. The dot is a symbol for multiplication. On some of the probes, you will see the dot used to stand for multiplication and on others, you will see an x, or ‘times’ sign. What is the correct answer to problem B, 7 times 9?*** [Wait for students to volunteer 63]. ***Yes, that’s correct. The correct answer is 63.***

Now we’ll take some time, about 20 seconds, so you can practice doing a Basic Facts probe. If you finish before I say ‘Stop’, please do NOT turn to the next page. Any questions?

Ready, begin. [Time for 20 seconds—watch clock or watch]

Stop. Please put your 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 of two Basic Facts probes. You will have 1 minute to work on this probe. Remember, your job is to answer as many problems correctly as you can in 1 minute. 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. If you get to the STOP page, remember to stop there.

****When I say begin, please turn past the sample/STOP page and start working. You will have 1 minute. Do your best work.***

Ready, (Pause) ***begin.*** [Start timer. Time for 1 minute.] When the timer goes off, say, ***Stop. Please put your pencils down. If you haven’t turned to the red STOP page, please do so. [Now we will do the second Basic Facts probe.]*** [REPEAT FROM * FOR SECOND PROBE.]

Now we will shift gears to a different type of probe.

Estimation Probe 1

A large pizza has 16 slices. 5 kids will share the pizza. About how many slices will each kid get? 0.3 3 30	22 x 59 is about 12 120 1,200	It takes 2 1/2 yards of fabric to make a costume for the play. Mom has 11 yards of fabric. About how many costumes can she make? 0.4 4 40	$8\overline{)555}$ is about 0.7 7 70	The gym shoes cost $82. They are on sale for 25% off. About how much will you save? $20 $200 $2,000
4 x 9.3 is about 0.36 3.6 36	Edward makes $4 per hour doing odd jobs. If he works 11 hours, about how much will he earn? $4 $40 $400	0.45 - 0.14 is about 0.3 3 30	Luis wants to buy 6 new books. If each one costs $12, about how much will Luis pay for the 6? $0.60 $6 $60	$8\overline{)0.19}$ is about 0.02 0.2 2
Christine's car went 300 miles on 11 gallons of gas. About how many miles per gallon did the car go? 27 270 2,700	97.7 - 21.4 is about 0.8 8 80	There are 30 students in the class. Each student paid $3.50 for the bus. About how much money do they have for a bus? $10 $100 $1,000	73 - 18 is about 5 50 500	Joel earns $4.25 per hour stacking shelves at the grocery store. About how much will he earn in 20 hours? $0.80 $8 $80
78 + 17 is about 10 100 1,000	The car's gas tank holds 14 gallons. You just pumped in 11.75 gallons to make the tank full. About how many gallons were already in the tank? 2 20 200	219 + 876 is about 10 100 1,000	For her birthday, Sue received $19 from Grandpa and $32 from Aunt Sue. About how much did she receive from these two people? $5 $50 $500	$82\overline{)71.3}$ about 1 10 100

Research Institute Progress Monitoring (RIPM)
2005-2006
Middle School Measures Study

Estimation
Standard Directions

Set timer to 3 minutes.

"***The second type of probe you will be doing is called an Estimation probe.*** [Hold up booklet with Estimation sample page showing.] ***This sample page shows some examples of the types of problems on the probe. Here's how you complete this probe. Read the problem and estimate the answer. DO NOT take time to calculate an exact answer. Look at the three choices at the bottom of the box and decide which one is the best estimate for the correct answer. If you're not sure about an answer, it is better to skip the problem than to take a wild guess. If you do too much guessing, you will lose points on this probe. Now look at Problem A. You should read the problem, 94 and one tenth minus 29 and four tenths, and estimate the answer. The answer is about 60, so best estimate is 60. Now look at problem B. What is the correct answer to Problem B?*** [Wait for students to volunteer *20*]. ***Yes, that's correct. 20 is about half of 42.***

Now we'll take a minute so you can practice doing an estimation probe. Remember that on some of the probes, you will see the dot used to stand for multiplication and on others, you will see an x, or 'times' sign. When you take this probe, please start here [point to upper left corner of 8 sample problems at the bottom of the page] ***and move across each row from left to right*** [demonstrate]. ***When I say begin, please work on the 8 practice problems at the bottom of the sample page. If you finish before I say 'Stop', please do NOT turn to the next page. Any questions?***

Ready, begin. [Time for 1 minute—watch clock or watch]

Stop. Please put your 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 of two Estimation probes. You will have *3* ***minutes to work on this probe. Remember, your job is to answer as many problems correctly as you can in 3 minute. You DO NOT need to figure out an exact answer. Just use mental math to choose the best estimate. Remember that it is better to skip a problem if you're not sure of the answer than to take a wild guess. 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. If you get to the STOP page, remember to stop there.***

****When I say begin, please turn past the sample/STOP page and start working. You will have 3 minutes. Do your best work.***

Ready, (Pause) ***begin*****.** [Start timer. Time for 3 minutes.] When the timer goes off, say, ***Stop. Please put your pencils down. If you haven't turned to the red STOP page, please do so. [Now we will do the second Estimation probe.]*** [REPEAT FROM * FOR SECOND PROBE.]

That is the end of the probes for today. THANK YOU for your hard work!

Complex Quantity Discrimination Form A

50 + 9 ☐ 9 + 50	18 ☐ 2 • 9	67 ☐ 76	.3 ☐ .03
56 ☐ 7 • 8	.5 ☐ .2	50 • 7 ☐ 5 • 70	15 ☐ 15 ÷ 3
32 ÷ 8 ☐ 32 ÷ 5	60 + 2 ☐ 20 + 6	60 • 4 ☐ 6 • 40	10 ☐ 10 – 7
70 + 3 ☐ 72 + 3	20 – 8 ☐ 20 – 9	85 ☐ 80	79 ☐ 70 + 9
53 ☐ 35	.4 ☐ .3	10 – 2 ☐ 13 – 2	1/6 ☐ 1/2
70 – 1 ☐ 71 – 1	50 ☐ 50 + 1	29 ☐ 20	26 ☐ 2 + 6
1/9 ☐ 1/5	90 ☐ 91 – 1	90 ☐ 90 + 2	20 ☐ 20 – 8
9 ☐ 90	48 ÷ 6 ☐ 48 ÷ 9	70 • 80 ☐ 70 • 8	80 + 4 ☐ 40 + 8
3 • 20 ☐ 3 • 2	15 ☐ 15 ÷ 5	74 ☐ 7 + 4	65 ☐ 60 + 5
1 • 90 ☐ 1 • 9	.9 ☐ .09	30 ☐ 34 – 4	80 – 9 ☐ 80 – 4
3 ☐ 30	40 • 50 ☐ 40 • 5	10 + 9 ☐ 9 + 10	60 + 4 ☐ 61 + 4

Project RIPM, Iowa State University

Research Institute Progress Monitoring (RIPM)
2005-2006
Middle School Measures Study

Complex Quantity Discrimination
Standard Directions

Set timer to 1 minute.

"The next type of probe you will be doing is called a Complex Quantity Discrimination probe. [Hold up booklet with Complex Quantity Discrimination sample page showing.] ***This sample page shows some examples of the types of problems on the probe. For each problem, you will look at the two quantities and then write greater than, less than, or equal to in the box.*** [As you say each term, write the corresponding symbol on the board: >, <, =] ***You can see that problem A has been answered. 60 plus 5 is less than 61 plus 5, so the correct answer is less than. Now look at problem B. The two quantities are 60 times 5 and 5 times 60. The dot between the two numbers is a symbol for multiplication. On some of the probes, you will see the dot used to stand for multiplication and on others, you will see an x, or 'times' sign. What is the correct answer to problem B?*** [Wait for students to volunteer *equal to*]. ***Yes, that's correct. The correct answer is equal to, so you would write an equal sign in the box.***

Now we'll take some time, about 20 seconds, so you can practice doing a Complex Quantity Discrimination probe. If you finish before I say 'Stop', please do NOT turn to the next page. Any questions?

Ready, begin. [Time for 20 seconds—watch clock or watch]

Stop. Please put your 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 of three Complex Quantity Discrimination probes. You will have 1 minute to work on this probe. Remember, your job is to answer as many problems correctly as you can in 1 minute. 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. If you get to the STOP page, remember to stop there.

****When I say begin, please turn past the sample/STOP page and start working. You will have 1 minute. Do your best work.***

Ready, (Pause) ***begin.*** [Start timer. Time for 1 minute.] When the timer goes off, say, ***Stop. Please put your pencils down. If you haven't turned to the red STOP page, please do so. [Now we will do the second/last Complex Quantity***

Discrimination probe.] [REPEAT FROM * FOR SECOND AND THIRD PROBES.]

Now we will shift gears to the last type of probe.

Missing Number Form A

37, 40, 43, ___	11, 22, 44, ___	43, ___, 45, 46	12, ___, 24, 30
30, 35, 40, ___	32, 28, ___, 20	21, 16, 11, ___	50, 60, ___, 80
48, 40, 32, ___	25, 20, 15, ___	24, ___, 12, 6	35, ___, 37, 38
18, 16, 14, ___	19, 20, 21, ___	6, ___, 12, 15	8, 16, 32, ___
80, 79, ___, 77	56, 57, ___, 59	88, 44, 22, ___	27, ___, 21, 18
24, 12, 6, ___	36, ___, 54, 63	14, 16, 18, ___	56, 49, ___, 35
70, 60, ___, 40	14, 21, ___, 35	49, 45, 41, ___	28, 25, 22, ___
10, 9, 8, ___	43, 45, 47, ___	47, 57, 67, ___	38, 42, 46, ___
2, 4, 8, ___	28, ___, 26, 25	7, 12, 17, ___	80, 40, 20, ___
40, 48, 56, ___	12, 16, ___, 24	81, ___, 63, 54	48, 38, 28, ___
17, 15, 13, ___	56, 28, 14, ___	13, 26, 52, ___	65, 64, ___, 62

Project RIPM, Iowa State University

Research Institute Progress Monitoring (RIPM)
2005-2006
Middle School Measures Study

Missing Number
Standard Directions

Set timer to 1 minute.

*"**The first type of probe you will be doing is called a Missing Number probe.*** [Hold up booklet with Missing Number sample page showing.] ***This sample page shows some examples of the types of problems on the probe. For each problem, you need to look at the numbers and try to figure out the missing number in the pattern. You can see that problem A has been answered. 63, 81, and 90 are all multiples of 90, so the correct answer is 72. Now look at problem B. The numbers are 26, 29, and 32. What number would come next in this pattern?*** [Wait for students to volunteer 35]. ***Yes, that's correct. The correct answer is 35, because the pattern is increasing by 3, so you would write 35 on the line.***

Now we'll take some time, about 20 seconds, so you can practice doing a Missing Number probe. If you finish before I say 'Stop', please do NOT turn to the next page. Any questions?

Ready, begin. [Time for 20 seconds—watch clock or watch]

Stop. Please put your 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 of three Missing Number probes. You will have 1 minute to work on this probe. Remember, your job is to answer as many problems correctly as you can in 1 minute. 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. If you get to the STOP page, remember to stop there.

****When I say begin, please turn past the sample/STOP page and start working. You will have 1 minute. Do your best work.***

Ready, (Pause) ***begin.*** [Start timer. Time for 1 minute.] When the timer goes off, say, ***Stop. Please put your pencils down. If you haven't turned to the red STOP page, please do so. [Now we will do the second/last Missing Number probe.]*** [REPEAT FROM * FOR SECOND AND THIRD PROBES.]

Now we will shift gears to a different type of probe.

Appendix B

Teacher Rating Form

Teacher ____________________

Middle School Math Progress Monitoring Study
_________ Middle School
Fall 2005

Directions:	For each student, rate his or her **general proficiency in mathematics** 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.

Student	Mathematics Proficiency						
	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
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7