Research Institute on Progress Monitoring

TECHNICAL REPORT #13:

Technical Adequacy of Early Numeracy Indicators: Exploring Growth at Three Points in Time

Anne Foegen, Erica Lembke, 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 130 Kindergarten and Grade 1 students in a Midwestern school. Two forms of two different early numeracy measures (Quantity Discrimination and Missing Number) were administered in the fall, winter, and spring of the 2005-06 school year. In the fall, the use of group administration was contrasted with typical individual administration procedures for Grade 1 students. Criterion measures included teacher ratings (fall and spring) and a norm-referenced test of mathematics ability, the TEMA-3 (spring only). Alternate form reliability results were strong for the Quantity Discrimination measure at both grade levels and for the Missing Number measure in Kindergarten. Concurrent criterion validity correlation coefficients were in the moderate range for both measures, with the Quantity Discrimination measure having higher coefficients than Missing Number in Kindergarten and the reverse pattern for Grade 1. Predictive validity results were moderate for teacher ratings for both measures at both grade levels. For the TEMA-3, relations were moderate for Grade 1 students, but low for Kindergarten students. Growth data revealed effect sizes ranging from .93 to 1.88. Mean weekly slope values ranged from .14 (Grade 1, Missing Number) to .51 (Kindergarten, Quantity Discrimination). In general, slopes values were higher for the Quantity Discrimination measure and for Kindergarten students.

Technical Adequacy of Early Numeracy Indicators: Exploring Growth at Three Points in Time

Considerable research exists on the technical adequacy and implementation of early screening measures in the area of reading. In mathematics, however, work in this area is in its infancy. Given recent emphasis on universal screening and early intervention for students at-risk, educators have an urgent need for tools in mathematics that will support these efforts. While measures exist for students in the elementary grades (cf., Fuchs, Fuchs, & Hamlett, 1998, 1999; Marston, 1989), these tools focus on formal computational skills and are not well suited to students in Kindergarten or early Grade 1.

Initial studies in the development of early numeracy measures include work by Clarke and his colleagues, who have explored tasks involving oral counting, number identification, quantity discrimination, and missing numbers in a sequence (Clarke & Shinn, 2004; Chard, Clarke, Baker, Otterstedt, Braun, & Katz, 2005). The results of these studies supported the number identification, quantity discrimination, and missing number measures as having the strongest alternate-form reliability and criterion validity, with validity coefficients ranging from $r = .50$ to .69 in Kindergarten and from $r = .60$ to .79 in Grade 1.

VanDerHeyden, Witt, Naquin, and Noell (2001) also examined early math measures for screening preschool and Kindergarten students to identify those in need of academic intervention. Kindergarten tasks, which included circle number, write number, and draw circles, had criterion validity coefficients ranging from $r = .44$ to .61 with a mathematics achievement test. For 4-year olds, tasks such as choose number, number naming, counting objects, free counting, discrimination, and choosing shapes were correlated with the Brigance Screens (Brigance, 1999) and the Test of Early Mathematics Achievement-2 (TEMA-2; Ginsburg &

Baroody, 1990). The strongest correlations were demonstrated with the choosing number and discrimination tasks.

Lembke and Foegen (2005) examined the reliability and criterion validity of four types of early math screening measures, including Number Identification, Quantity Discrimination, Quantity Array, and Missing Number, in two studies conducted in Missouri and Iowa. Their studies included 119 Kindergarten and 118 Grade 1 students. Criterion measures included teacher ratings, the Mini Battery of Achievement (Woodcock, McGrew, & Werder, 1994), the Stanford Early Achievement Test (Psychological Corporation, 1996), and the Test of Early Mathematics Ability-3 (TEMA-3; Ginsburg & Baroody, 2003). Alternate-form reliability coefficients for single probes ranged from r = .72 to .89; test-retest coefficients ranged from r = .73 to .91. Criterion validity coefficients ranged from r = .36 to .71, with the majority of coefficients in the r = .50 to .60 range. Lembke and Foegen concluded that of the four measures investigated, the Number Identification, Quantity Discrimination, and Missing Number tasks demonstrated the strongest technical adequacy. Following this study, the results were used to inform revisions to the measures. In particular, the Missing Number measure was adapted to include a higher (80%) proportion of items using the standard counting sequence (count by 1s) than used in the initial study (70%). In addition, items requiring counting by 2s were eliminated, so the remaining 20% of the items consisted of counting by 5s or 10s. Thirty percent of the earlier measure had been comprised of items involving counting by 2s, 5s, or 10s. No revisions were made to the Quantity Discrimination measure.

The purpose of the present study is to replicate Lembke and Foegen's initial study by examining reliability and criterion validity of slightly revised measures and to extend their initial work by examining changes in student performance across multiple administrations of the measures within a single school year. We used the following research questions to guide our study:

1. What levels of alternate-form reliability are produced by the revised early numeracy indicators?
2. What levels of criterion validity are associated with the revised early numeracy indicators?
3. To what extent does group administration of the early numeracy indicators influence the reliability and validity of the measures?
4. To what extent does students' performance on the measures change across multiple administrations (fall, winter, spring) within a school year?

Method

Setting and Participants

The study was conducted in a small Midwestern district on the fringe of an urban community. The district had four schools within the district: a preschool to third grade school, a fourth/fifth grade school, a middle school, and a high school. During the 2005-06 school year, the district enrolled 1,457 students, 53% male, 92% white, 4.4 % Hispanic, and 3% other ethnicities. Nearly forty-three percent of the students in the district qualified for free and reduced lunch, and 17% of students were receiving special education services. All four Kindergarten and four Grade 1 teachers in the building were invited and consented to participate in the study. Parental consent was obtained for one hundred thirty students (55 K and 75 Grade 1 students). Table 1 displays the demographic characteristics of the student participants.

Measures

Early Mathematics Indicators. Two early mathematics measures were investigated in this study: Quantity Discrimination and Missing Number. The Quantity Discrimination task required students to name the greater of two numbers. The task consisted of 63 pairs of numbers; students had one minute to respond verbally by naming the larger number in each pair. Numerals from 0 to 20 were used to create the items. We scored the probe by counting the number of correct responses in one minute.

Table 1

Demographic Characteristics of Student Participants

	Kindergarten (n = 55)	Grade 1 (n = 75)	Total Percentage
Gender			
Male	30	39	53%
Female	25	36	47%
Ethnicity			
White	53	69	94%
Hispanic	2	4	5%
Other ethnicities	0	2	1.5%
Free/reduced lunch	22	30	40%
ESL Services	1	3	3%
Special Education Services	0	2	1.5%

To complete the Missing Number task, students were presented with a series of three numbers and one blank indicating a missing number in the sequence (the position of the blank varied). Students responded by verbally naming the missing number. All items used forward counting sequences that involved counting by 1s, 5s, or 10s. The task consisted of 63 items;

students had one minute to respond. We scored the probe by counting the number of correct responses in one minute. Samples of the measures, including construction guidelines and administration directions, are provided in Appendix A.

Two forms of each type of measure were used for the individual administration. The same two measures were used for each of the three data collection periods. We created two additional forms of each measure for use in the group administration portion of the study, which was only conducted in the fall.

Criterion measures. The criterion measures used in the study included teachers' ratings of their students' overall math proficiency, standardized test scores, a computation-based Curriculum-Based Measurement task, and a comprehensive assessment associated with the district mathematics curriculum. 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 1 to 7, with 1 representing low proficiency and 7 representing high proficiency. Teachers were asked to use the entire scale, not clustering students only in the middle or toward one end. All teachers completed student ratings in the fall and the spring, concurrent with the respective probe administration procedures. A copy of the teacher rating form is included in Appendix B.

The *Test of Early Mathematics Ability-3* (TEMA-3; Ginsburg & Baroody, 2003) was used to provide a norm-referenced assessment of each student's mathematics proficiency. The TEMA-3 is designed for children ages 3 to 8 and consists of 72 items in the domains of informal and formal mathematics. Informal mathematics items assess four areas: numbering skills, number-comparison facility, calculation skills, and understanding of concepts. The formal mathematics items assess numeral literacy, mastery of number facts, calculation skills, and understanding of concepts. Whereas the informal items typically rely on verbal and pictorial or concrete representations, the majority of the formal items use written representations with traditional symbolic formats. The test manual reports internal consistency reliability coefficients of $r = .94$ to .96, alternate-form reliability coefficients of $r = .93$ to .97, and test-retest reliability coefficients of $r = .82$ to .93. The authors provide evidence for the content validity of the measure in three ways. First, they offer qualitative evidence of content validity by reporting a detailed rationale for the selection of items in the informal and formal mathematics domains. Second, they report the quantitative item analysis statistics (item discrimination and item difficulty) used to select items included in the final versions of the TEMA-3. Finally, they report the results of differential item functioning analyses used to examine potential bias through the differential performance of students on the basis of race/ethnicity, gender, or other demographic variables.

Criterion validity evidence of the TEMA-3 is presented in terms of correlation coefficients relating TEMA-3 performance to scores on other standardized, norm-referenced measures of mathematics achievement. These coefficients ranged from $r = .54$ to .55 to $r = .91$ (Ginsburg & Baroody, 2003). The authors provide evidence of the measure's construct validity by demonstrating changes in mean performance by age at each of six age intervals. They report means by various groups (gender, ethnicity, low mathematics achievement) and note that only the low mathematics achievement mean was outside of the normal range. Finally, the authors note that item validity is documented through strong correlations between scores on individual items and the total scale score.

Trained project personnel individually administered the TEMA-3 to all participating students in the spring during the same time period when data were being gathered on the early numeracy measures. The TEMA-3 Math Ability Score, a standard score with a mean of 100 and a standard deviation of 15, was used in the analyses.

Two additional criterion measures were available for Grade 1 students. Given that a computation-based progress monitoring measure with known technical adequacy exists for Grade 1 students, we were interested in exploring relations between scores on that measure and students' scores on the early mathematics indicators. We group-administered two forms of the Monitoring Basic Skills Progress – Computation Grade 1 measures (MBSP-Comp; Fuchs, Hamlett, & Fuchs, 1998) to students in each participating classroom in the spring. The MBSP-Comp is comprised of 25 addition and subtraction problems, most of which are single-digit number combinations. Students worked on the task for 2 minutes; we scored the probes by counting the number of digits correct. The second additional criterion measure used in Grade 1 was an end-of-course assessment for the district's curriculum program (*Growing with Mathematics*; Irons, 2003). The assessment included 59 problems that addressed topics such as numeration, basic computation, money, time, measurement, geometry, and problem solving. Teachers had administered the assessment under untimed conditions as part of their regular classroom activities at the conclusion of the school year. Percent correct scores were recorded for the analyses.

Procedures

Teachers explained the study to students and sent consent letters home. Teachers then collected consent forms and provided students with a pencil as an incentive to return the forms. Early mathematics measures and the MBSP-Comp were administered to all students in each participating classroom. The data reported here are only for those students for whom parental consent was provided to participate in the research component of the project.

Students participated in three rounds of data collection spread across the school year. Fall data were gathered in late October and early November, winter data in early February, and spring data in late April. During each data collection period, two forms of each task were individually administered, with each data collection session lasting approximately seven minutes per child. Administration of the tasks took place at desks or tables in the hallways outside of the students' classrooms. The order of the tasks was counterbalanced across classes. For both types of probes, the administration of the two forms was preceded by a brief introduction to the measure and three sample problems to insure that the student understood the task. If students were absent, "make-up" data collection sessions were attempted for one week. If students could not be assessed in this time frame, no data were recorded, but the students were included in subsequent (e.g., winter and spring) rounds of data collection using standard procedures.

Also in the fall, we examined whether group administration of the measures might represent a potential option for increasing the efficiency of screening large numbers of students. In Kindergarten classes, we administered two forms of the Quantity Discrimination measure to each class as a large group. Students responded to the probes by circling the greater number in each pair. As with the individual administration, we provided students with a brief introduction to the task and an opportunity to practice three sample problems. The order of tasks for group administration was counterbalanced across classrooms. We opted not to administer the Missing Number task to Kindergarten students, as doing so would require a level of fluency in writing numerals that we felt represented an unreasonable expectation. We administered two forms of both the Quantity Discrimination and Missing Number measures to Grade 1 classes. No make-up data were gathered for the group-administered measures.

In the spring, the TEMA-3 (Ginsburg & Baroody, 2003) was individually administered to the students for whom parental consent had been obtained. As with the early mathematics probes, the TEMA-3 was administered at tables or desks in the hallways outside the participating classrooms. Also in the spring, the first author administered the MBSP-Comp to full classes of students in each of the four participating Grade 1 classrooms. Standard administration directions (obtained from the manual for the program) and timing requirements were used. No make-up sessions were provided for the MBSP-Comp. A summary of the data collection activities is presented in Table 2.

Table 2

Data Collection Activities

	Kindergarten			Grade 1		
	F	W	S	F	W	S
Early Numeracy Indicators						
Individual: Quantity Discrimination	X	X	X	X	X	X
Individual: Missing Number	X	X	X	X	X	X
Group: Quantity Discrimination	X			X		
Group: Missing Number				X		
Criterion Measures						
Teacher Ratings	X		X	X		X
MBSP-Computation			X			X
TEMA-3			X			X
District End of Course Test						X

Note. F = Fall, W = Winter, S = Spring.

Trained data collectors gathered all of the data. Each data collector participated in a small-group training session lasting approximately one hour. Data collectors who administered the TEMA-3 attended an additional training session lasting approximately one hour. A graduate student who had been trained by the first author delivered this training session using scripted materials. An overview of the study was provided, and then the administration of each task was modeled for the data collectors. The data collectors practiced administering each of the tasks and then administered each task to a peer while the trainer observed and completed an 11-item fidelity checklist. Overall, the average percentage of accuracy on the fidelity checklist (computed by dividing the number of items completed correctly by 11) was 99% with a range of 91% to 100%.

Project staff completed all of the scoring and data entry. We rescored a minimum of ten percent of the probes in each of the three rounds to assess inter-scorer agreement. More probes were scored in the fall to assure that newly trained project staff members were completing the scoring accurately. 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 interscorer agreement. We present the scoring accuracy results in Table 3. These data support the conclusion that the early mathematics measures can be scored with a high degree of consistency and accuracy, with mean accuracy above 99% for both measures at each time point.

Table 3

Mean Agreement, Range and Number of Probes Examined for Interscorer Agreement

	Quantity Discrimination			Missing Number		
	Mean Agreement	Range	# Probes Rescored	Mean Agreement	Range	# Probes Rescored
Fall	.996	.88 – 1.0	134	.998	.92 – 1.0	104
Winter	1.0	1.0	32	1.0	1.0	32
Spring	.996	.83 – 1.0	46	.997	.88 – 1.0	46

Scoring and Data Analyses

Data analyses were conducted using number correct scores for each of the early mathematics measures. Alternate-form reliability was computed by correlating scores from the two individual probes of each type. For the criterion measures, teacher ratings were standardized by classroom and the resulting *z*-scores were used in the analyses. For the TEMA-3, we entered the standardized math ability score (mean = 100, SD = 15) in the analyses. We computed the number of correct digits on each of the two forms of the MBSP-Comp and used the average of these two scores for all analyses. The end of course test score was the percent of items answered correctly on the assessment provided by the curriculum publisher. We examined criterion validity by correlating the mean of the two probe forms for each measure with the criterion measures. We obtained rough estimates of growth first by examining effect sizes and then by computing OLS regression slopes across each student’s three data points.

Results

The results section begins with descriptive statistics for all the study measures. We next move to analyses specific to each of the research questions. Means and standard deviations for each of the individually administered early 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 Individually Administered Early Mathematics Screening Measures

Kindergarten							
Measure	Date	Measure	n	Min	Max	M	SD
Missing Number	Fall	Form A	51	0	21	6.61	5.77
		Form B	52	0	19	6.92	5.48
	Winter	Form A	55	0	21	10.95	4.56
		Form B	55	0	21	11.53	4.82
	Spring	Form A	53	3	21	13.36	4.84
		Form B	53	4	27	13.66	4.36
Quantity Discrimination	Fall	Form A	51	0	37	14.20	9.39
		Form B	52	0	32	12.02	9.21
	Winter	Form A	55	3	41	19.55	10.35
		Form B	55	0	42	19.53	10.18
	Spring	Form A	53	4	44	25.30	9.58
		Form B	53	2	45	22.72	9.48
Grade 1							
Measure	Date	Measure	n	Min	Max	M	SD
Missing Number	Fall	Form A	73	2	27	16.49	4.66
		Form B	73	4	25	16.48	4.47
	Winter	Form A	66	6	31	19.14	4.79
		Form B	66	8	32	18.58	4.51
	Spring	Form A	72	7	30	19.10	4.84
		Form B	72	10	33	20.06	4.42
Quantity Discrimination	Fall	Form A	73	6	49	31.45	9.20
		Form B	73	4	53	30.11	10.21
	Winter	Form A	66	11	59	36.85	8.62
		Form B	66	6	57	36.29	8.78
	Spring	Form A	72	19	60	40.11	7.71
		Form B	72	13	59	38.67	8.50

In reviewing the data in Table 4, we considered the nature of the distributions produced on each of the measures. We were particularly interested in any floor effects and the size of the standard deviations. Some Kindergarten students obtained scores of zero during both the fall and winter data collection periods; this occurred for both of the measures. As we examined the data more closely, we found that this occurrence was more frequent for the Missing Number measure (for which 10 to 13 students obtained scores of zero of the two forms in the fall) than for the Quantity Discrimination measure, for which only two students obtained scores of zero on each of the forms in the fall. By the time of the winter data collection, no more than one of the 55 Kindergarten students obtained a score of zero on any of the probe forms. Every Grade 1 student was able to respond with more than one correct response to each probe during all administration periods.

We examined the standard deviations produced for each measure and found that the Quantity Discrimination measure produced a much wider distribution of scores in both grade levels than did the Missing Number measure. In addition, students' mean scores on both measures and at both grade levels increased from one administration period to the next.

The descriptive statistics for group-administered data are presented in Table 5. Recall that these measures were administered in the Fall. In all cases, the mean scores for the group –administered, paper/pencil measures were lower than the scores obtained using individual administration with verbal responses from students. Although some scores of 0 were obtained, this was infrequent (only 1 or 2 students obtaining a 0 on any form). The group-administered data did not reveal the same differences in standard deviations (with substantially larger deviations for the Quantity Discrimination measure as compared to the Missing Number measure) we observed in the individually administered data. This may be due to variability associated with responding in writing, rather than verbally.

Table 5

Descriptive Statistics for Group-Administered Data

Measure	Grade	Measure	n	Min	Max	M	SD
Quantity Discrimination	K	Form A	51	0	21	10.16	5.74
		Form B	51	0	21	11.47	6.70
Missing Number	1	Form A	75	0	21	12.15	4.27
		Form B	75	0	21	11.25	4.21
Quantity Discrimination	1	Form A	75	10	42	23.96	6.21
		Form B	75	8	42	24.39	6.25

We present descriptive statistics for the criterion measures in Table 6. The *z*-scores for teacher ratings were calculated by classroom to control for variability in teachers' application of the rating scale. Teacher ratings for spring were lower than those for fall. Scores on the MBSP-Comp revealed that no students obtained scores of zero on the measure. The mean score on the TEMA-3 was just slightly above the normative mean of 100, while the standard deviation was somewhat smaller.

Table 6

Descriptive Statistics for Criterion Variables

Measure	n	Min	Max	M	SD
Kindergarten					
Teacher Rating *z* Score, Fall	51	-1.94	1.64	0	.97
Teacher Rating *z* Score, Spring	53	-2.17	1.21	0	.97
TEMA-3 Math Ability Score	51	67.00	123	101.43	11.69
Grade 1					
Teacher Rating *z* Score, Fall	73	-2.01	2.08	0	.98
Teacher Rating *z* Score, Spring	73	-2.08	1.41	0	.98
MBSP Computation	68	3.00	28.5	16.62	5.53
End of Grade Test	69	62.00	100	91.36	8.06
TEMA-3 Math Ability Score	73	78.00	133	104.66	13.72

In Table 7, we report the intercorrelations between the individual- and group-administered early numeracy indicators. We were particularly interested in the correlations between the individual and group-administered measures given concurrently in the fall. For the Kindergarten students, there was a low level of correspondence between the scores students obtained on the measures when they were administered in a group setting and those obtained when the measures were administered individually. The relations were stronger for Grade 1 students, but the strength of the relation was not sufficient to suggest that group administration could be used in place of individual administration.

Intercorrelations between the criterion variables are presented in Table 8. We observed positive relations between all of the criterion measures, but the size of the coefficients was not so strong that we were concerned about an unreasonable degree of overlap between the measures.

Table 7

Intercorrelations Between Early Numeracy Indicators

Kindergarten	Quantity Discrimination				Missing Number			
	Group	Individual			Group	Individual		
	F	F	W	S	F	F	W	S
Quantity Discrimination								
Group: Fall	--							
Individual: Fall	.29*	--						
Winter	.42**	.84**	--					
Spring	.31*	.77**	.90**	--				
Missing Number								
Individual: Fall	.09	.66**	.56**	.49**		--		
Winter	.34*	.49**	.70**	.62**		.58	--	
Spring	.28*	.43**	.67**	.73**		.44	.79**	--

Grade 1	Quantity Discrimination				Missing Number			
	Group	Individual			Group	Individual		
	F	F	W	S	F	F	W	S
Quantity Discrimination								
Group: Fall	--							
Individual: Fall	.65**	--						
Winter	.71**	.89**	--					
Spring	.64**	.79**	.86**	--				
Missing Number								
Group: Fall	.61**	.60**	.61**	.53**	--			
Individual: Fall	.61**	.76**	.66**	.65**	.64**	--		
Winter	.59**	.76**	.77**	.73**	.63**	.78**	--	
Spring	.56**	.64**	.70**	.74**	.63**	.71**	.82**	--

Note. ** $p < .01$. * $p < .05$. Average score from two forms used for each measure.

Table 8

Intercorrelations Between the Criterion Measures

		1	2	3	4	5
1	Teacher Rating, Fall	--				
2	Teacher Rating, Spring	.81**	--			
3	TEMA-3 Math Ability Score	.32*	.53**	--		

		1	2	3	4	5
1	Teacher Rating, Fall	--				
2	Teacher Rating, Spring	.75**	--			
3	TEMA-3 Math Ability Score	.68**	.71**			
4	MBSP Computation, Average	.51**	.53**	.58**	--	
5	End of Grade Test	.48**	.57**	.49**	.49**	--

Note. ** $p < .01$. * $p < .05$.

Research Question 1: Alternate Form Reliability of the Early Numeracy Indicators

Alternate form reliability results for the individually administered measures are presented in Table 9. In general, the coefficients for Quantity Discrimination were at or near the $r = .80$ benchmark, while those for Missing Number were somewhat lower. Students at both grade levels were more consistent in their performance across forms on the Quantity Discrimination measure than on the Missing Number measure. No consistent patterns were observed favoring the reliability of the measures at one grade level over another.

Table 9

Alternate Form Reliability Coefficients for Individually Administered Early Numeracy Indicators

	Quantity Discrimination				Missing Number			
	Kindergarten		Grade 1		Kindergarten		Grade 1	
	N	*r*	N	*r*	N	*r*	N	*r*
Fall	51	.91	73	.91	51	.90	73	.75
Winter	55	.74	66	.84	55	.69	66	.73
Spring	53	.93	72	.91	53	.66	72	.75

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

Research Question 2: Criterion Validity of the Early Numeracy Measures

Because of differences in the criterion measures across grade levels, the results of our criterion validity analyses are reported separately by grade level. In Tables 10 and 11, we report the concurrent validity coefficients for Kindergarten and Grade 1 students, respectively. In general, concurrent validity coefficients were in the moderate range. Different patterns of results emerged for Kindergarten and Grade 1 students. At the Kindergarten level, Quantity Discrimination coefficients were consistently higher than those obtained for Missing Number. In the Grade 1 data, the coefficients for Missing Number were similar or slightly higher than those for Quantity Discrimination. Stronger relations were obtained with teacher ratings and the TEMA-3 than with the MBSP-Comp and the end-of-grade assessment.

Table 10

Concurrent Validity Coefficients for Kindergarten Students

	Fall				Spring			
	QD		MN		QD		MN	
Fall Criterion Measure	N	*r*	N	*r*	N	*r*	N	*r*
Teacher Ratings	49	.69	49	.55				
Spring Criterion Measures								
Teacher Ratings					53	.78	53	.63
TEMA-3					51	.59	51	.42

Note. All correlations significant, $p < .01$. QD = Quantity Discrimination; MN = Missing Number.

Table 11

Concurrent Validity Coefficients for Grade 1 Students

	Fall				Spring			
	QD		MN		QD		MN	
Fall Criterion Measure	N	*r*	N	*r*	N	*r*	N	*r*
Teacher Ratings	73	.57	73	.62				
Spring Criterion Measures								
Teacher Ratings					72	.59	71	.60
TEMA-3					72	.55	71	.55
MBSP-Comp					67	.49	67	.48
End-of-Grade Test					68	.34	68	.35

Note. All correlations significant, $p < .01$. QD = Quantity Discrimination; MN = Missing Number.

We next examined the predictive validity of the measures by computing correlations between the fall probe scores and the spring criterion measures. These results for both grade levels are presented in Table 12. As with the concurrent validity results for Kindergarten, fall administration scores from the Quantity Discrimination measure were more strongly related to spring criterion measures than were scores from the Missing Number measure. In Grade 1, the results varied by criterion measure, with some criterion variables more closely related to the Quantity Discrimination measure (TEMA-3, MBSP-Comp) and other more closely related to the Missing Number measure (end-of-grade test). Predictive validity coefficients for Kindergarten students were much higher for teacher ratings than for the TEMA-3. In Grade 1, coefficients for all measures were in the moderate range, with the exception of the end-of-grade test, for which correlations were low to moderate.

Table 12

Predictive Validity Coefficients for Kindergarten and Grade 1 Students

	Kindergarten				Grade 1			
	QD		MN		QD		MN	
Spring Criterion Measure	N	*r*	N	*r*	N	*r*	N	*r*
Teacher Ratings	50	.64	50	.59	71	.62	71	.61
TEMA-3	48	.42	48	.34	71	.67	71	.63
MBSP-Comp					66	.62	66	.53
End-of-Grade Test					67	.37	67	.53

Note. All correlations significant, $p < .05$. QD = Quantity Discrimination; MN = Missing Number.

Research Question 3: Effects of Group Administration on Technical Adequacy of the Early Numeracy Measures

We next examined the group-administered data to determine the alternate-form reliability and criterion validity of the measures. In Table 13 we present the results of the alternate-form reliability analyses; results of the correlations between group-administered early mathematics measures and selected criterion variables are presented in Table 14. The reliability coefficients were far below the standard benchmark of .80. These results call into question the viability of using group administration to obtain reliable student scores. The criterion validity coefficients in

Table 14 were promising for Grade 1 students (all in the $r = .5$ to .6 range), but were low for Kindergarten students. These findings will also have to be considered in light of the reliability and criterion validity results for the individually administered measures.

Table 13

Alternate Form Reliability of Group-Administered Early Numeracy Indicators

	Kindergarten		Grade 1	
	N	*r*	N	*r*
Quantity Discrimination	51	.68	75	.66
Missing Number			75	.64

Note. All correlations significant at p < .01.

Table 14

Criterion Validity Results for Group- Administered Early Numeracy Indicators

	Quantity Discrimination				Missing Number	
	Kindergarten		Grade 1		Grade 1	
	N	*r*	N	*r*	N	*r*
Fall Teacher Rating	49	.30*	73	.66**	73	.57**
Spring Teacher Rating	50	.33*	73	.61**	73	.58**
Spring TEMA-3	48	.04	73	.60**	73	.51**

Note. * = $p < .05$; ** = $p < .01$.

Research Question 4: Growth on the Early Numeracy Indicators

Our final research question examined the extent to which students' scores on the early numeracy indicators 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 in standard deviation units. As the results in Table 15 indicate, effect sizes exceeding one standard deviation were obtained for both measures at both grade levels. The largest effect sizes were obtained for the Missing Number measure at both Kindergarten and Grade 1. Although these findings seem to favor the Missing Number measure, readers must also keep in mind the impact of standard deviation size in interpreting these results. For the Missing Number measure, standard deviations were in the 4 to 5 point range, suggesting growth from spring to fall of about 8 to 10 points on the measure. For the Quantity Discrimination measure, the effects sizes were smaller, but the standard deviations were larger (in the 7 to 10 point range), so anticipated changes in actual probe scores for the Quantity Discrimination measure would be in the 7 to 14 point range.

Table 15

Standardized Within-Grade Growth on Early Numeracy Indicators

	Kindergarten		Grade 1	
Measures	ES	*n*	ES	*n*
Quantity Discrimination	1.40	*50*	0.93	*70*
Missing Number	1.88	*50*	1.46	*70*

Note. ES denotes standardized effect size: a difference score for each measure (average score in spring minus average score in fall) was computed for each student and the mean difference was expressed in standard deviations.

To further examine the changes in student performance from fall to spring testing, we computed linear regression slope values for each student that reflected the rate of change in scores across the three testing periods. We divided the resulting values by 11 (the approximate number of weeks between each administration period) to estimate the amount of expected growth per week on each measure. The results of these analyses are presented in Table 16.

Table 16

Weekly Slope Values by Grade Level

	Kindergarten			Grade 1		
Measures	Mean (SD)	Min	Max	Mean (SD)	Min	Max
Quantity Discrimination	.51 (.28)	.02	1.20	.38 (.28)	-.48	1.02
Missing Number	.32 (.23)	-.07	.86	.14 (.15)	-.27	.55

At both grade levels, student scores increased more rapidly on the Quantity Discrimination measure than on the Missing Number measure. Rates of growth were larger for students in Kindergarten than for students in Grade 1.

Discussion and Future Research

The results of this study extend the work of Lembke and Foegen (2005), who documented the technical adequacy of early numeracy measures for use as static indicators of student proficiency. Our results were comparable with regard to alternate-form reliability. For Kindergarten students, both measures had single form alternate-form reliability levels exceeding $r = .90$. At Grade 1, the alternate-form reliability of the Quantity Discrimination measure exceeded $r = .90$, but was .75 for the Missing Number measure.

With regard to criterion validity, our concurrent validity coefficients for teacher ratings were in the $r = .55$ to .69 range, nearly identical to Lembke and Foegen's (2005) results, which ranged from $r = .58$ to .69. In the present study, we used a different norm-referenced achievement criterion measure (TEMA-3, rather than the Mini Battery of Achievement), but again obtained quite similar concurrent validity coefficients ($r = .42$ to .59 in the present study vs. $r = .38$ to .55 in the earlier study).

We extended work in the area of early numeracy measures by examining predictive validity and growth over time. Our predictive validity results were in the high moderate range (r = .59 to .64) for teacher ratings, with no patterns favoring one measure over another at either grade level. Predictive validity coefficients for the TEMA-3 produced a different pattern of results, with low relations (coefficients of r = .34 and .42) between Kindergarteners' fall scores on the early numeracy measures and their spring scores on the TEMA-3. The predictive validity of the measures for Grade 1 was much stronger, ranging from r = .63 to .67.

Effect sizes reflecting student growth from fall to spring revealed increases of *ES* = .93 to 1.88, with larger *ES*s for the Missing Number measure and higher *ES* for Kindergarten students than for Grade 1 students. When we examined student growth by computing individual slopes across the three data points, we found that mean weekly slope values were higher for the Quantity Discrimination measure at both grade levels. Slopes for Kindergarten students were higher than those for Grade 1 students.

The results of this study provide additional evidence of the potential of the early numeracy measures for use as screening tools and progress monitoring measures in Kindergarten and Grade 1. Future research should examine the use of the measures for more frequent progress monitoring and explore the sensitivity of the measures in detecting responsiveness to an intervention program.

References

Chard, D. J., Clarke, B., Baker, S., Otterstedt, J., Braun, D., & Katz, R. (2005). Using measures of number sense to screen for difficulties in mathematics: Preliminary findings. *Assessment for Effective Intervention, 30*(2), 3-14.

Clarke, B., & Shinn, M. R. (2004). A preliminary investigation into the identification and development of early mathematics curriculum-based measurement. *School Psychology Review, 33*, 234-248.

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

Ginsburg, H. P., & Baroody, A.J. (1990). *Test of Early Mathematics Ability, 2nd edition.* Austin, TX: Pro-Ed, Incorporated.

Ginsburg, H. P. & Baroody, A.J. (2003). *Test of Early Mathematics Ability*, 3rd edition. Austin, TX: PRO-ED.

Irons, C. (2003). *Growing with mathematics*. Guilford, CT: Wright Group/McGraw Hill.

Lembke, E., & Foegen, A. (2005). *Identifying indicators of early mathematics proficiency in Kindergarten and Grade 1* (Technical Report 6). Minneapolis, MN: University of Minnesota, College of Education and Human Development, Research Institute on Progress Monitoring.

Marston, D. (1989). Curriculum based measurement: What it is and why do it? In M. R. Shinn (Ed.), *Curriculum-based measurement: Assessing special children* (pp. 18-78). New York: Guilford.

Stanford Early Achievement Test (9th edition, 1996). Psychological Corporation. Harcourt Publishing.

VanDerHeyden, A. M., Witt, J. C., Haquin, G., Noell, G. (2001). The reliability and validity of curriculum-based measurement readiness probes for Kindergarten students. *School Psychology Review*, *30*, 363-382.

Woodcock, R. W., McGrew, K. S., & Werder, J. K. (1994). *Mini-Battery of Achievement*. Itasca, IL: Riverside Publishing.

Appendix A

Early Mathematics Measures

Quantity Discrimination

Construction Procedures

Administration Procedures

Sample Quantity Discrimination Measure

Missing Number

Construction Procedures

Administration Procedures

Sample Missing Number Measure

Quantity Discrimination

Construction Procedures

- Used number sets 0-10 and 0-20
- Randomly selected either 0-10 or 0-20
- From the selected number set, selected two numbers for each problem
- If the next two random numbers are identical (i.e., next two numbers are both 2’s) eliminate one, and move to the next number.

Short Directions for Quantity Discrimination:

1. Place the student copy in front of the student.

2. Place the examiner copy on a clipboard and position so the student cannot see what the examiner records.

3. Say these specific directions to the student:

> **"Look at the paper in front of you. In each row there are some boxes with numbers in them."** (Point to the first set of boxes in the top row). **"I want you to tell me the number that is bigger."**

4. Correct Response:

> **"Good. 7 is bigger than 1.**" (Point to the second set of boxes in the top row.)

Incorrect Response:

> **"The number that is bigger is 7. You should have said 7 because 7 is bigger than 1."** (Point to the second set of boxes in the top row.)

5. Continue with the other example(s). After the examples, turn to the first page of the student copy of the probe.

6. Say to the student:

> **"When I say begin, I want you to tell me which number is bigger. Start here and go across the page** (demonstrate by pointing). **Try each one. If you come to one that you don't know, I'll tell you what to do. Are there any questions? Put your finger on the first one. Ready, begin."**

7. Start your stopwatch. If the student fails to attempt (does not give the answer to the first problem) after 3 seconds, tell the student to

> **"Try the next one."**

8. For at least the first 2 to 3 rows of problems, you may need to prompt the student by pointing to the next box and saying

> **"Tell me which number is bigger."**

9. On the administrator copy, write the number that the student says in the blank next to each problem number.

10. The maximum time for each item is 3 seconds. If a student does not provide an answer within 3 seconds, tell the student to

> **"Try the next one."**

11. If the student comes to the end of the page, turn the page to the next page of problems.

12. At the end of 1 minute, draw a line under the last item completed and say

> **"Stop."**

13. Repeat these directions for probe 2.

Scoring Rules

Rule 1: If a student correctly identifies the number score the item as correct.

Rule 2: If the student states any number other than the item number score the item as incorrect.

Rule 3: If a student hesitates or struggles with a problem for 3 seconds tell the student to "try the next one" and score the item as incorrect.

Rule 4: If a student skips a problem, score the problem as incorrect.

Rule 5: If a student skips an entire row, mark each problem in the row as incorrect.

Examples

1	7

6	2

8	0

Quantity discrimination, page 1—student copy

5	2	7	1	8	3
1	18	8	10	7	8
16	8	9	1	10	7
2	6	8	3	9	4
12	5	9	15	10	8
0	14	0	6	8	10
15	14	6	1	5	1

Quantity discrimination, page 2—student copy

8	4	5	9	16	1
1	9	8	5	19	0
1	0	4	5	5	10
18	10	13	14	0	2
3	10	7	0	8	7
7	9	12	2	9	1
11	18	17	18	9	17

Quantity discrimination, page 3—student copy

8	1	9	15	15	18
4	18	9	2	3	7
8	6	15	6	6	8
9	6	10	1	2	8
12	7	5	8	5	3
4	20	6	0	15	17
2	9	6	5	1	5

Quantity Discrimination—Administrator copy

Student:________________________Date:______________Number correct:_____
Directions: Write the number that the student says in the blank.

1. __________ (5)

2. __________ (7)

3. __________ (8)

4. __________ (18)

5. __________ (10)

6. __________ (8)

7. __________ (16)

8. __________ (9)

9. __________ (10)

10. _________ (6)

11. _________ (8)

12. _________ (9)

13. _________ (12)

14. _________ (15)

15. _________ (10)

16. _________ (14)

17. _________ (6)

18. _________ (10)

19. _________ (15)

20. _________ (6)

21. _________ (5)

22. _________ (8)

23. _________ (9)

24. _________ (16)

25. _________ (9)

26. _________ (8)

27. _________ (19)

28. _________ (1)

29. _________ (5)

30. _________ (10)

31. _________ (18)

32. _________ (14)

33. _________ (2)

34. _________ (10)

35. _________ (7)

36. _________ (8)

37. _________ (9)

38. _________ (12)

39. _________ (9)

40. _________ (18)

41. _________ (18)

42. _________ (17)

43. _________ (8)

44. _________ (15)

45. _________ (18)

46. _________ (18)

47. _________ (9)

48. _________ (7)

49. _________ (8)

50. _________ (15)

51. _________ (8)

52. _________ (9)

53. _________ (10)

54. _________ (8)

55. _________ (12)

56. _________ (8)

57. _________ (5)

58. _________ (20)

59. _________ (6)

60. _________ (17)

61. _________ (9)

62. _________ (6)

63. _________ (5)

Missing Number

Construction Procedures

- Use forward counting sequence
- 80% of the problems are counting by 1's and 20% are counting by 5's and 10's
- Problem type is randomly selected
- Number that the problem starts with is randomly selected (0-7 for count by 1's, 5-35 for count by 5's, and 10-70 for count by 10's)
- Includes problems counting by 1-digit from 0-10, by 5's to 50, and by 10's to 100.
- For counting by 1-digit, the blank varies
- For counting by 5's and 10's, the blank is at the end.
- 3 numbers are given, with a blank in-between or at the end (student completes the pattern by stating or writing the 4^{th} number)

Short Directions for Missing Number, Individual Administration:

1. Place the student copy in front of the student.

2. Place the examiner copy on a clipboard and position so the student cannot see what the examiner records.

3. Say these specific directions to the student:

> **"Look at the paper in front of you. Each box has three numbers and a blank."** (Point to the first box). **"What number goes in the blank?"**

4. Correct Response:

> **"Good. The number is 3**." (Point to the second box.)

Incorrect Response:

> **"The number that goes in the blank is 3**. **You should have said 3 because 3 comes after 2 (0, 1, 2, 3)."** (Point to the second box.)

5. Continue with the other example(s). After the examples, turn to the first page of the student copy of the probe.

6. Say to the student:

> **"When I say begin, I want you to tell me what number goes in the blank in each box. Start here and go across the page** (demonstrate by pointing). **Try each one. If you come to one that you don't know, I'll tell you what to do. Are there any questions? Put your finger on the first one. Ready, begin."**

7. Start your stopwatch. If the student fails to attempt the first problem, after 3 seconds tell the student to

> **"Try the next one."**

8. For at least the first 2 to 3 rows of problems, you may need to prompt the student by pointing to the next box and saying

> **"Tell me the number that goes in the blank."**

9. On the administrator copy, write the number that the student says in the blank next to each problem number.

10. The maximum time for each item is 3 seconds. If a student does not provide an answer within 3 seconds, tell the student to

"Try the next one."

11. If the student comes to the end of the page, turn the page to the next page of problems.

12. At the end of 1 minute, draw a line under the last item completed on the administrator page and say

"Stop."

13. Repeat these directions for probe 2.

Scoring Rules

Rule 1: If a student correctly identifies the number score the item as correct.

Rule 2: If the student states any number other than the item number score the item as incorrect.

Rule 3: If a student hesitates or struggles with a problem for 3 seconds tell the student to "try the next one" and score the item as incorrect.

Rule 4: If a student skips a problem, score the problem as incorrect.

Rule 5: If a student skips an entire row, mark each problem in the row as incorrect.

Missing Number

Example 1

0 1 2 ___	1 ___ 3 4	5 10 15 ___

Missing Number, page 1—Student copy

___ 8 9 10	___ 7 8 9	4 ___ 6 7
30 40 50 ___	1 2 3 ___	4 5 ___ 7
___ 3 4 5	4 5 6 ___	7 8 ___ 10
3 ___ 5 6	7 8 ___ 10	6 7 8 ___
10 15 20 ___	6 7 ___ 9	1 2 ___ 4
3 4 5 ___	4 ___ 6 7	5 6 7 ___
0 1 ___ 3	___ 1 2 3	5 ___ 7 8

Missing Number, page 2—student copy

4 5 6 ___	35 40 45 ___	5 6 ___ 8
6 7 8 ___	3 ___ 5 6	0 ___ 2 3
7 8 9 ___	10 20 30 ___	30 40 50 ___
1 ___ 3 4	1 2 3 ___	___ 4 5 6
7 8 ___ 10	5 6 7 ___	___ 2 3 4
4 ___ 6 7	2 3 ___ 5	3 4 5 ___
6 7 8 ___	1 ___ 3 4	___ 7 8 9

Missing Number, page 3—student copy

___ 7 8 9	2 ___ 4 5	1 2 3 ___
0 1 2 ___	5 6 ___ 8	3 4 ___ 6
___ 1 2 3	0 ___ 2 3	2 3 4 ___
1 ___ 3 4	60 70 80 ___	0 1 2 ___
5 10 15 ___	5 6 ___ 8	10 15 20 ___
6 7 8 ___	0 ___ 2 3	0 1 2 ___
2 3 ___ 5	___ 5 6 7	6 ___ 8 9

Missing Number—Administrator copy

Student:________________________Date:______________Number correct:_____
Directions: Write the number that the student says in the blank.

1. __________ (7)

2. __________ (6)

3. __________ (5)

4. __________ (60)

5. __________ (4)

6. __________ (6)

7. __________ (2)

8. __________ (7)

9. __________ (9)

10. __________ (4)

11. __________ (9)

12. __________ (9)

13. __________ (25)

14. __________ (8)

15. __________ (3)

16. __________ (6)

17. __________ (5)

18. __________ (8)

19. __________ (2)

20. __________ (0)

21. __________ (6)

22. __________ (7)

23. __________ (50)

24. __________ (7)

25. __________ (9)

26. __________ (4)

27. __________ (1)

28. __________ (10)

29. __________ (40)

30. __________ (60)

31. __________ (2)

32. __________ (4)

33. __________ (3)

34. __________ (9)

35. __________ (8)

36. __________ (1)

37. __________ (5)

38. __________ (4)

39. __________ (6)

40. __________ (9)

41. __________ (2)

42. __________ (6)

43. __________ (6)

44. __________ (3)

45. __________ (4)

46. __________ (3)

47. __________ (7)

48. __________ (5)

49. __________ (0)

50. __________ (1)

51. __________ (5)

52. __________ (2)

53. __________ (90)

54. __________ (3)

55. __________ (20)

56. __________ (7)

57. __________ (25)

58. __________ (9)

59. __________ (1)

60. __________ (3)

61. __________ (4)

62. __________ (4)

63. __________ (7)

Appendix B

Teacher Rating Form

Early Math Measures Study
Teacher Rating of Students' Math Proficiency

Teacher Name: ____

Directions: Please list the names of each of the students participating in the project below. Think about each student in the context of peers of the same age/grade level. Please rate each student's ***general proficiency in math*** relative to other students in the same grade level. Students who have very low levels of math proficiency compared to their peers should be rated a 1. Those who have very high levels should be rated a 7. Thank you for your help!

Student Name	**1** *(least proficient)*	**2**	**3**	**4**	**5**	**6**	**7** *(most proficient)*
	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