TECHNICAL REPORT #28:

Exploring the Use of Early Numeracy Indicators for Monitoring Progress in Two Intervention Contexts

Linda Impecoven-Lind, Jeannette Olson, and Anne Foegen

RIPM Year 4: 2006 – 2007

Dates of Study: September 2006 – May 2007

August 2009

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

Two of the Early Numeracy Indicators developed by Lembke and Foegen (2005) were used as progress monitoring measures with small groups of first grade students. This report describes the alternate-form reliability and sensitivity to growth of the Quantity Discrimination and Missing Number measures when used to monitor progress. In addition, the efficacy of two intervention approaches was explored. On the whole, much lower levels of alternate-form reliability were achieved when the measures were used for progress monitoring than when they were used for screening. Both measures appear to be sensitive to growth, with students showing significant improvement on the Quantity Discrimination measure. We also examined the growth rates of students who participated in two types of small group interventions. The first was an intervention developed by classroom teachers. The second was the 3-Tier Mathematics Model materials developed at the University of Texas (Bryant et al., 2008). Although students in both groups demonstrated improvement with supplemental instruction, no significant differences in growth were found between the two interventions.

Exploring the Use of Early Numeracy Indicators for Monitoring Progress

in Two Intervention Contexts: 2006-2007

The reliability and criterion validity of two Early Numeracy Indicators developed by Lembke and Foegen (2005) for use as benchmarking tools have been examined in previous studies (Lembke & Foegen, 2005; Foegen, Lembke, Klein, Lind, & Jiban, 2006). In this study, two of the Early Numeracy Indicators were used as progress monitoring measures with small groups of first grade students who were struggling with beginning math concepts. This report examines the reliability and sensitivity to growth of these assessments. In addition, we examine the effects of two kinds of mathematics interventions with these students.

Research Questions

The following research questions guided the data analysis:

1. What levels of alternate-form reliability are demonstrated when the Early Numeracy Indicators are used as progress monitoring measures?
2. To what extent do the two progress monitoring measures reflect changes in student performance?
3. Did the use of different mathematics interventions result in different mean scores or growth rates on the Early Numeracy Indicators?

In the sections that follow, the method and results for the progress monitoring research questions (1 and 2) are presented first, followed by the method and results for the intervention research question (3).

Method: Progress Monitoring

Setting and Participants

The study was conducted in a small Midwestern district on the fringe of an urban community. The district included four schools: two elementary schools, a middle school, and a high school. During the 2006-2007 school year, the district enrolled 1,424 students, with 52.9 percent being male, 93.3 percent white, 3.7 percent Hispanic, and 3 percent other ethnicities. Nearly 44 percent of the students qualified for free or reduced price lunch, and approximately 17 percent of students were receiving special education services.

Although the Early Numeracy Indicators (ENIs) were used as benchmarking assessments for all kindergarten and first grade students, the progress monitoring measures were only used with selected groups of first grade students from the four first grade classes at the elementary school. The four first grade classes included a total of 86 students, of whom 59.3 percent were males, 90.7 percent were white, and 9.3 percent were Hispanic. Seven percent of the first grade students were receiving special education services.

Gathering the early numeracy data was a part of the school's typical practices and ongoing commitment to making data driven decisions; therefore, the study was deemed exempt from typical Human Subjects procedures. Individual consent was not required for students' participation in the data collection efforts.

After all of the fall benchmarking scores had been entered in a database, the Project Coordinator computed each student's percentile rank and prepared class lists of test results. Student scores that fell into the 90^{th} percentile and above, 80^{th} to 89^{th} percentile, 11^{th} through 20^{th} percentile, and below the 10^{th} percentile were color coded on these lists. The Project Coordinator and the Principal Investigator met with the principal and each of the teachers to review the data.

The first grade teachers assigned seventeen students to the intervention groups during the fall meetings, all of whom scored below the 20th percentile on at least one of the ENI measures and were not already receiving supplemental or special education services in mathematics. Table 1 presents the scores and percentiles for the students assigned to the intervention groups. Eleven of the intervention students had at least one ENI score that fell at or below the 10th percentile, and six had at least one score between the 11th and 20th percentile.

Measures

Early Numeracy Progress Monitoring Measures. Two ENIs (Quantity Discrimination and Missing Number) were used monitor the progress of the students in the intervention groups. (See Appendix A for samples of one page of each type of measure.) Quantity Discrimination required students to name the greater of two numbers. The task consisted of 42 pairs of numbers (as compared to 63 items on the benchmarking Quantity Discrimination task). Students responded verbally by naming the number with the greatest value in each pair. Numerals from 0 to 20 were used to create the items. Numbers were randomly selected using a random number generator. For Missing Number, 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). 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 42 items (as compared to 63 items on the benchmarking Missing Number task). As with the Quantity Discrimination measure, a random number generator was used to determine the beginning number as well as the position of the blank. Twelve forms of each of the two kinds of assessments were developed. Individual student progress monitoring booklets were prepared that included the directions, twelve alternate forms of the measure, and a blank graph for charting the student's score on each administration.

Table 1

Fall Benchmarking Scores and Percentiles

		Number Identification		Quantity Discrimination		Missing Number	
Student	Teacher	Score	Percentile Range	Score	Percentile Range	Score	Percentile Range
1	1	6	< 10	3	< 10	1.5	< 10
2	1	10.5	11-20	18	11-20	8.5	<10
3	1	10	11-20	13.5	11-20	17.5	21-79
4	1	32.5	21-79	15	11-20	12.5	21-79
5	2	21	21-79	14	11-20	16.5	21-79
6	2	10.5	11-20	18	11-20	13.5	21-79
7	2	15	11-20	10.5	< 10	13	21-79
8	3	15.5	11-20	5	< 10	10.5	11-20
9	3	12.5	11-20	12.5	< 10	12	11-20
10	3	7.5	< 10	5	< 10	10	< 10
11	3	21.5	21-79	18	11-20	10.5	11-20
12	3	12.5	11-20	12	< 10	10.5	11-20
13	3	13.5	11-20	17.5	11-20	1	< 10
14	4	14.5	11-20	9	< 10	10	< 10
15	4	9	< 10	23	21-79	13.5	21-79
16	4	13.5	11-20	24	21-79	13	21-79
17	4	8.5	< 10	15.5	11-20	10.5	11-20

Procedures

Classroom teachers gathered all of the progress monitoring data. Near the beginning of the school year, each teacher participated in a small-group training session lasting approximately one hour that was delivered by the Project Coordinator and the Principal Investigator. During the training session an overview of the benchmarking process was provided, then the project coordinator modeled how to administer each of the three measures and the procedure for scoring the benchmarking measures. The procedure for administering and scoring the progress monitoring measures was the same as the one used for benchmarking; therefore, teachers did not receive any additional training for using the progress monitoring measures.

Both tasks were individually administered using a printed testing booklet. Students were given one minute to verbally respond to as many items as they could during this time period. If a student hesitated for three seconds on a particular item, he/she was prompted to "Try the next one." Teachers recorded student responses in a printed booklet and later scored the measures by counting the number of correct responses. Scores were plotted on the graph included in the progress monitoring booklet. At a later date, a member of the research team entered all of the scores into a spreadsheet for the data analysis.

Students participated in nineteen weeks of data collection spread across the academic year. The measures were administered once a week (generally on Friday). Quantity Discrimination probes were used to monitor student progress during the first 14 weeks of the interventions. Beginning with Week 13, the teachers returned to the first probe and cycled back through the set of twelve Quantity Discrimination progress monitoring measures. At Week 15, teachers expressed concerns that students were improving in their Quantity Discrimination proficiency, but that their scores were not representative of their global proficiency in mathematics. As a result, we decided

to monitor progress using the Quantity Discrimination and Missing Number tasks during alternate weeks.

Data Analyses

Data analyses were conducted using number correct scores for the two kinds of progress monitoring probes. Alternate-form reliability was computed by correlating scores from adjacent administrations of the same measure. To determine the measures' sensitivity to growth, we used ordinary least squares regression to calculate the slope of each student's scores on the two measures. The obtained slope values were calculated to reflect the amount of weekly progress a student achieved.

Results: Progress Monitoring

Descriptive statistics for all of the study measures are reported first. These are followed by alternate-form reliability coefficients and student growth data for the two ENIs that were used as progress monitoring assessments. Means and standard deviations for the Quantity Discrimination progress monitoring measures are presented in Table 2, as are data for the intervention students on the Fall, Winter, and Spring screening measures. Table 3 includes similar information for the Missing Number measures.

As we considered the data in Tables 2 and 3 we looked at the distributions for each of the progress monitoring measures. We were most interested in floor or ceiling effects and the progression of the means over time. There were no scores of zero for any of the weeks that the two measures were used. However, there were at least three weeks when one or more students earned the maximum score of 42 on the Quantity Discrimination assessment. This happened as early as Week 9 of the intervention. None of the scores on the Missing Number tasks were close to the total points possible.

Table 2

Descriptive Statistics for Quantity Discrimination Progress Monitoring Measures

Measure	Week	Form	*n*	Min	Max	*M*	*SD*
Quantity		Fall Screen	17	3	24	13.74	5.97
Discrimination	1	1	16	2	30	19.50	8.40
	2	2	15	8	34	21.33	8.13
	3	3	16	13	31	22.75	5.25
	4	4	16	14	33	23.62	5.28
	5	5	16	18	33	27.56	4.49
	6	6	16	19	39	29.12	5.06
	7	7	15	21	39	28.80	4.71
	8	8	15	17	39	29.93	6.38
		Winter Screen	16	22.5	36	31.28	4.14
	9	9	14	24	42	33.21	5.86
	10	10	14	21	39	31.50	5.35
	11	11	14	24	42	33.93	5.84
	12	12	9	24	35	30.67	3.61
	13	13	9	29	40	35.22	3.99
	14	14	8	28	40	34.38	3.74
	16	15	10	26	41	35.70	5.76
	18	1	11	20	42	32.91	6.35
		Spring Screen	15	16.5	41.5	31.67	6.52

Table 3

Descriptive Statistics for Missing Number Progress Monitoring Measures

Measure	Week	Form	*n*	Min	Max	*M*	*SD*
Missing		Fall Screen	17	1	16.5	10.79	4.21
Number		Winter Screen	16	9.0	19.0	15.43	2.60
	15	1	15	11	21	16.80	3.03
	17	2	15	12	24	18.67	3.13
	19	3	11	13	24	17.36	3.61
		Spring Screen	15	7.5	23	17.38	4.40

Examining the progression of mean scores on the two measures, we found students' scores increased over the course of the year on both measures. Mean scores on Quantity Discrimination

improved during each of the first six weeks of the study; however, during the remaining twelve weeks, the scores showed much more variability often going up one week and down the next. Mean scores on Missing Number did increase during the first two weeks it was used, but also dropped as the academic year was ending.

Research Question 1. What levels of alternate-form reliability are demonstrated when the Early Numeracy Indicators are used as progress monitoring measures?

The alternate-form reliability correlations between scores from each type of probe for each pair of subsequent administrations are reported in Table 4. The statistically significant correlations ranged from .60 to .87 for Quantity Discrimination. None of the three Missing Number coefficients for alternate-form reliability were statistically significant.

Table 4

Alternate-form Reliability

Measure	Forms	*N*	*r*	*p*
Quantity Discrimination	QD 1 and QD 2	15	.79	.00
	QD 2 and QD 3	15	.87	.00
	QD 3 and QD 4	16	.70	.01
	QD 4 and QD 5	16	.40	.12
	QD 5 and QD 6	16	.60	.01
	QD 6 and QD 7	15	.61	.02
	QD 7 and QD 8	14	.23	.43
	QD 8 and QD9	14	.67	.01
	QD 9 and QD 10	14	.65	.01
	QD 10 and QD 11	14	.63	.02
	QD 11 and QD 12	9	.70	.04
	QD 12 and QD 1	9	.22	.56
	QD 1 and QD 2	8	.65	.08
	QD 2 and QD 3	8	-.02	.96
	QD 3 and QD 4	10	.48	.16
Missing Number	MN 1 and MN 2	15	.40	.14
	MN 2 and MN 3	11	.45	.17

Research Question 2: To what extent do the two progress monitoring measures reflect changes in student performance?

Table 5 includes the data on student growth rates on the two progress monitoring measures examined in this study. Quantity Discrimination progress monitoring measures were used to assess students at 16 points across 19 weeks of the school year. The mean rate of improvement for all of the intervention students was 1.21 correct responses each week. Student growth rates ranged from 0.07 to 3.00 correct responses per week.

The Missing Number measures were only used for three occasions across a five-week period toward the end of the study. As a result, the growth rates reported here should be viewed with great caution. Six students showed negative slopes on the Missing Number measure. While the overall mean growth rate was .55 correct responses per week, a substantial level of variability was evident in students' growth rates, which ranged from -1.00 to 4.00 correct responses per week.

To determine if student performance improved significantly from the beginning to the end of the progress monitoring period, we used the Wilcoxon Signed Ranks Test. After examining graphs of each student's Quantity Discrimination scores, we used the mean of the first two scores and the mean of the last two scores for this comparison. Students' scores improved significantly from the start to the end of the progress monitoring period on the Quantity Discrimination measure ($z = -3.23$, $p < .01$). As noted earlier, the Missing Number measures were only used during three weeks of the study. Students' scores on the first and last administration of Missing Number (a span of five weeks of school) were not significantly different ($z = -0.51$, $p > .61$).

Table 5

Weekly Growth Rates

		Slopes	
Student	Teacher	Quantity Discrimination	Missing Number
1	1	1.48	-0.75
2	1	1.60	-0.50
3	1	1.10	-1.00
4	1	1.24	-0.25
5	2	0.69	4.00
6	2	3.00	2.50
7	2	0.89	3.00
8	3	1.08	1.00
9	3	1.67	0.50
10	3	2.36	-1.00
11	3	0.07	0.00
12	3	0.31	-0.50
13	3	0.50	NA
14	4	2.69	NA
15	4	0.57	0.25
16	4	0.69	0.75
17	4	0.56	0.25

Note. Students 13 and 14 moved before the Missing Number measures were used for progress monitoring.

Method: Intervention

In addition to determining how well the Quantity Discrimination and Missing Number measures worked for monitoring student progress, we also examined the relative efficacy of two intervention approaches that were used with the small groups of students that were included in this study. Two teachers used "teacher-developed" (TD) interventions, while two teachers used the 3-Tier Mathematics Model (3TMM) materials that were developed at the University of Texas (Bryant et al., 2008). The principal selected which teachers would implement which interventions.

Materials

The TD intervention was the existing approach to intervention at the school participating in this study. This approach relied on teachers' knowledge from classroom instruction regarding the types of skills and concepts on which identified students needed additional reinforcement. Using this information, they selected supplemental practice activities from the existing curriculum or they developed their own materials to provide students with additional practice.

The 3TMM intervention used a direct instruction approach with scripted lessons that addressed number sense, place value/relationships of 10, problem solving, and addition and subtraction combinations. Most lessons included modeling, guided practice, individual practice, and suggestions for corrective feedback. These lessons were designed to be used for 15 to 20 minutes during four days of each week.

Procedures

All of the students who were included in the intervention groups received mathematics instruction in their general education classrooms. They also received supplemental instruction from their classroom teachers. The students in the classes taught by Teacher 1 and Teacher 2 participated in the TD intervention, while the students in the classes taught by Teacher 3 and

Teacher 4 participated in the 3TMM intervention. Both groups of students were monitored using the progress monitoring measures described earlier in this report.

Data Analysis of Intervention Data

We chose to use the nonparametric Mann-Whitney *U* test and the Wilcoxon Signed Ranks Test to examine the intervention data because the sample was small and included only low performing students. There were seven students in the TD intervention group and ten students in the 3TMM intervention group.

Research Question 3: Did the use of different mathematics interventions result in different mean scores or growth rates on the Early Numeracy Indicators?

First, we wanted to determine whether or not the intervention groups were equivalent prior to beginning the interventions. Table 6 includes the mean scores for the three fall screening measures for the TD intervention group and the 3TMM intervention group and the results of the Mann-Whitney Test, which revealed that the groups were not significantly different on any of these measures.

Table 6

Comparison of Fall Benchmarking Scores

	TD			3TMM				
Fall Screening	*n*	*M*	*SD*	*n*	*M*	*SD*	*U*	*p*
Number Identification	7	15.07	9.03	10	12.85	4.06	34.00	.96
Quantity Discrimination	7	13.14	5.19	10	14.15	6.70	33.00	.89
Missing Number	7	11.71	5.26	10	10.15	3.45	22.00	.23

Next, we examined the growth rates exhibited by the two groups of students on the two progress monitoring measures to see if these rates were significant. We used the Wilcoxon Signed Ranks Test using the same combination of scores that we used when addressing Research Question 2 (means of the first two and last two Quantity Discrimination scores, and the first and last Missing Number scores), but we split the file by group for this analysis. The results, presented in Table 7, show that students in both intervention groups demonstrated significant improvement from the beginning to the end of the progress monitoring period on Quantity Discrimination. Also consistent with the results for Research Question 2, we found no significant differences for Missing Number. We next used the Mann-Whitney *U* Test to examine differences in the growth rates (i.e., slopes) on the two measures for the two intervention groups**.** The results were non-significant for Quantity Discrimination ($U = 27.00$, $p = .96$) and for Missing Number ($U = 21.00$, $p = .19$).

Table 7

Comparison of Improvement from Beginning to End of the Intervention Period

	TD					3TMM				
Measure	*n*	Pre *M*	Post *M*	*z*	*p*	*n*	Pre *M*	Post *M*	*z*	*p*
Quantity Discrimination	7	14.67	33.79	-2.20	.03	10	24.00	35.13	-2.38	.02
Missing Number	7	15.14	14.25	-1.84	.07	8	18.38	19.14	-1.05	.29

Finally, we wanted to see if one of the interventions led to higher scores on the spring benchmarking measures than the other intervention. The results presented in Table 8 reveal no significant differences between the spring benchmarking scores for the two groups.

Table 8

Comparison of Spring Benchmarking Scores

	TD			3TMM				
Spring Screening	*n*	*M*	*SD*	*n*	*M*	*SD*	*U*	*p*
Number Identification	7	43.64	3.77	8	42.63	9.51	24.00	69
Quantity Discrimination	7	30.57	4.89	8	32.63	7.89	20.00	.40
Missing Number	7	15.64	4.37	8	18.88	4.10	17.00	.23

Discussion

We were surprised by the alternate-form reliability correlations obtained when the measures were used for progress monitoring. We anticipated coefficients close to the .80 level, similar to the correlations obtained when we used the Quantity Discrimination and Missing Number indicators as benchmarking measures (Impecoven-Lind, Olson, & Foegen, 2009). However, the alternate-form reliability correlations reached this level for only two of the fifteen comparisons for Quantity Discrimination and for neither of the Missing Number comparisons. There are several possible reasons for the discrepancies we found. First, whereas the screening measures were administered by evaluators away from the classroom setting, the classroom teachers administered the measures inside the classroom as other students were doing other activities. This may have been more distracting for some students. In addition, no fidelity checks were conducted with the classroom teachers; it is possible that teachers did not adhere carefully to the standardized administration procedures. Finally, it is possible that student motivation varied from week to week when the measures were routinely administered for progress monitoring purposes.

We were encouraged that the two Early Numeracy Indicators that we chose to monitor student progress appear to be sensitive to student growth. The rates of growth for the Quantity Discrimination measure were strong even though they were used only with students struggling with mathematics concepts. Although Missing Number was used for just three of the last five weeks of the study, most students' scores did increase during this short time frame; however, future studies should examine the use of the Missing Number measure for monitoring progress over a longer time period.

Improvements in student scores resulted from both intervention approaches. As we reported earlier, there were no significant differences between each group's rates of growth or spring screening scores. We should note that the 3TMM lessons were still being refined during the 2006-2007 academic year; therefore, it would be worthwhile to continue to explore the efficacy of the two approaches to intervention in subsequent studies.

References

Bryant, D., Bryant, B., Gersten, R., Scammacca, N., Funk, C., Winter, A., Shi, M., & Pool, C. (2008). The effects of tier 2 intervention on the mathematics performance of first-grade students who are at risk for mathematics difficulties. *Learning Disability Quarterly, 31,* 47-63.

Foegen, A., Lembke, E., Klein, K., Lind, L., & Jiban, C. (2006). *Technical adequacy of early numeracy measures: Exploring growth at three points in time* (Technical Report 13). Minneapolis, MN: University of Minnesota, College of Education and Human Development, Research Institute on Progress Monitoring.

Impecoven-Lind, L., Olson, J., & Foegen, A. (2009). *Early numeracy indicator screening data: 2006-2007* (Technical Report 22). Minneapolis, MN: University of Minnesota, College of Education and Human Development, Research Institute on Progress Monitoring.

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.

Appendix A

Early Numeracy Progress Monitoring Measures

Quantity Discrimination

Sample Quantity Discrimination Measure Page

Missing Number

Sample Missing Number Measure Page

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

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