TECHNICAL REPORT #29:

Exploring the Use of Early Numeracy Indicators for Monitoring Progress in Two Intervention Contexts: 2007-08

Jeannette Olson and Anne Foegen

RIPM Year 5: 2007 – 2008

Dates of Study: September 2007 – May 2008

September 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 (Lembke & Foegen, 2005; Olson, Foegen, & Singamaneni, 2009; Nagarkar, Hamptom, Lembke, & Whitaker, 2009) 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 Mixed Numeracy measures when used to monitor progress. In addition, the efficacy of two intervention approaches was explored. This was the second year that Quantity Discrimination was used for progress monitoring, and it had much higher alternate-form reliability correlations than it had in the previous academic year. Mixed Numeracy was used as a benchmarking and progress monitoring measure for the first time in the current study. On the whole, the alternate-form reliability correlation coefficients improved over the course of the year as students and teachers became more familiar with the format and nature of the measures. Both measures appear to be sensitive to growth, with students showing significant improvement on both measures. 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: 2007-2008

This study extends research examining the use of Early Numeracy Indicators (ENIs) developed by Lembke and Foegen (Lembke & Foegen, 2005; Olson, Foegen, & Singamaneni, 2009; Nagarkar, Hamptom, Lembke, & Whitaker, 2009) for progress monitoring. In this study, two of the ENIs were used as progress monitoring measures with small groups of first grade students who were struggling with beginning math concepts. This study examined the reliability and sensitivity to growth of these assessments. In addition, we examined the effects of two kinds of mathematics intervention with low-performing students.

Research Questions

The following research questions guided the data analysis:

1. What levels of alternate-form reliability are demonstrated when the ENIs 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 an elementary school (grades Pre-K-3) in a small Midwestern school district on the fringe of an urban community. The school district is composed of four schools. There is one Pre-K through third grade elementary school, one fourth and fifth grade elementary school, one middle school with grades six through eight, and one high school. During the 2007-2008 school year, the district enrolled 1,464 students, with 46.4 percent being female, 90 percent white, 5.5 percent Hispanic, 2.7 percent African American, 1.5 percent Asian, and 0.3 percent Native American. Nearly 49 percent of the students qualified for free or reduced lunch, and 1.8 percent were identified as English Language Learners. Approximately 18 percent of the students received special education services.

Although the ENIs were used as benchmarking assessments for all kindergarten and first grade students, the progress monitoring measures were only used with selected groups of students from the four first grade classes at the elementary school. The four classes included a total of 108 students, of whom 46.3 percent were females, 88.9 percent were white, 6.5 percent were Hispanic, 3.7 percent were African American, and 0.9 percent were Native American. During the 2007-2008 academic year, 17.6% of the first grade students received special education services and 2.8% were classified as English Language Learners

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 needed 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 90th percentile and above, 80th to 89th percentile, 11th through 20th percentile, and below the 10th 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 thirteen students to the intervention groups during the fall meetings. In all but one case, the teachers chose students who scored below the 20th percentile on at least one of the four benchmarking measures and were not already receiving supplemental or special education services in mathematics. Table 1 presents the scores and percentile ranges for the students who were assigned to the intervention groups. Ten of the intervention students had at least one ENI score that fell below the 10th percentile, and three students had a least one score between the 11th and 20th percentile. There was one student who scored in the 21st through 79th percentile range on each of the measures. He was included in an intervention group because his teacher was concerned about his classroom mathematics performance.

Table 1

Fall Benchmarking Scores and Percentiles

		Number Identification		Quantity Discrimination		Missing Number		Mixed Numeracy	
Student	Teacher	Score	Percentile Range	Score	Percentile Range	Score	Percentile Range	Score	Percentile Range
1	1	23	21-79	22	21-79	8.5	< 10	16.5	11-20
2	1	12	< 10	15	11-20	10.5	21-79	17	11-20
3	1	10.5	< 10	15.5	11-20	8.5	< 10	14	< 10
4	2	15.5	11-20	13.5	< 10	6.5	< 10	15.5	11-20
5	2	17	11-20	23	21-79	9.5	11-20	19.5	21-79
6	3	12	< 10	7.5	< 10	2.5	< 10	11.5	< 10
7	3	10	< 10	9.5	< 10	3.5	< 10	9.0	< 10
8	3	19.5	21-79	15	11-20	8.5	< 10	14	< 10
9	3	10.5	< 10	9	11-20	8	< 10	11	< 10
10	4	10.5	< 10	17.5	< 10	13.5	21-79	18	21-79
11	4	29	21-79	19.5	21-79	13.5	21-79	18.5	21-79
12	4	24.5	21-79	15	11-20	9.5	11-20	18.5	21-79
13	4	25	21-79	18	11-20	13	21-79	18	21-79

Measures

Early Numeracy Progress Monitoring Measures. Two ENIs (Quantity Discrimination and Mixed Numeracy) 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 by using a random number generator. For the Mixed Numeracy measures, students were presented with items that were similar to the three other screening measures (see Olson, Foegen, & Singamaneni, 2009, for more information about these measures). This measure began with a row of four Number Identification items (name the number), followed by a row of four Quantity Discrimination items (name the greater of the two numbers), and then a row of four Missing Number items (name the numeral that is missing from a sequence of four numbers). This sequence repeated for a total of 84 items. Fifteen forms of each of the two kinds of tasks were developed. Individual progress monitoring booklets were prepared for each measure that included the directions, fifteen alternate forms, and a blank graph for charting the student's score from each administration.

Procedures

Classroom teachers gathered all of the progress monitoring data. All of the teachers had participated in the progress monitoring process during the 2006-2007 academic year; therefore, the Project Coordinator reviewed the administration procedures with the teachers instead of conducting a special training session.

The measures were administered once a week (generally on Friday) from November through April. Teachers alternated between the Quantity Discrimination and Mixed Numeracy measures, giving one during one week, and the other during the next week. Both tasks were individually administered using printed testing materials. 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 the appropriate booklet and later scored the measures by counting the number of correct responses. Scores were plotted on a graph that was 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 data analysis.

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. We used the Wilcoxon Signed Ranks Test to determine if students’ growth from the beginning to the end of the progress monitoring period was statistically significant.

Results: Progress Monitoring

Descriptive statistics for all of the study measures are reported first. These are followed by alternate-form reliability statistics and the student growth data for the two ENIs that were used as progress monitoring assessments. Means and standard deviations for the Quantity Discrimination and Mixed Numeracy 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 the same information for the Mixed Numeracy measures. Teachers started using the measures during different weeks of the school year and they had some weeks when they were not able to administer the progress monitoring measures; therefore, some teachers were using different forms of the measures during most of the weeks of the study.

Table 2

Descriptive Statistics for the Quantity Discrimination Progress Monitoring Measures

Measure	Week of School	Measure	*n*	Min	Max	*M*	*SD*
Quantity Discrimination	8	Fall Screen	13	7.5	23	15.39	4.76
	12	1	2	32	33	32.50	.71
	14-15	1 or 2	13	14	35	26.00	7.26
	16-17	2 or 3	9	21	36	30.33	6.08
	18	4	2	31	32	31.50	.71
	19	3 or 5	13	17	35	25.85	6.73
	21	4 or 6	13	17	38	31.15	6.27
	23	5 or 7	13	16	42	30.85	8.05
	25	Winter Screen	13	14	38.5	29.35	7.24
	26	6 or 8	12	19	42	30.92	7.68
	27-28	7 or 9	13	21	42	36.31	6.91
	29-30	8 or 10	12	23	42	36.08	5.92
	31-32	9 or 11	13	22	42	36.15	6.39
	33-35	10 or 12	9	25	42	36.78	6.59
	34	Spring Screen	13	22.5	43.5	32.50	7.87
	36	11 or 13	9	25	42	37.33	6.83

Table 3

Descriptive Statistics for the Mixed Numeracy Progress Monitoring Measures

Measure	Week of School	Measure	*n*	Min	Max	*M*	*SD*
Mixed Numeracy	8	Fall Screen	13	9	19.5	15.46	3.33
	13	1	13	14	27	19.62	3.75
	15-16	2	13	14	28	21.62	4.50
	17-18	3	12	17	28	22.33	3.87
	19	4	2	18	24	21.00	4.24
	20	4 or 5	13	20	30	25.31	3.38
	22	5 or 6	13	18	34	26.92	4.66
	24	6 or 7	13	20	35	28.00	4.73
	25	Winter Screen	13	16	33	24.39	5.10
	26-27	7 or 8	13	21	32	26.38	3.25
	28-29	8 or 9	13	23	42	31.08	4.86
	30-31	9 or 10	13	25	42	32.46	5.43
	32-33	10 or 11	12	22	37	30.75	4.98
	34	Spring Screen	13	22.5	41	31.00	5.45
	35-36	10, 11, or 12	8	30	44	34.88	5.69

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 means over time. There were no scores of zero for any of the weeks that the two measures were used. However, there were seven weeks when one or more students earned the

maximum score of 42 on the Quantity Discrimination assessment. One student earned the maximum score before the Winter screening period. The Mixed Numeracy progress monitoring measures have 84 items, and the highest score ever achieved by the students in the intervention groups was 44, so there were no ceiling effects with this assessment.

As we examined the progression of mean scores on the two measures, we found that the scores on both measures increased over time. There was less consistent growth on the Quantity Discrimination measure than on the Mixed Numeracy measure with several weeks where there was little or no change in the mean score. The limited growth on the Quantity Discrimination measure was probably related to the number of students who were earning the maximum score for several weeks. When we examined the pattern for the Mixed Numeracy mean scores, we found some growth for eight of the ten week-to-week progress monitoring comparisons.

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 task for each pair of subsequent administrations are reported in Tables 4 and 5. All of the correlations for Quantity Discrimination were statistically significant, with the coefficients ranging from .58 to .88.

Table 4

Alternate-form Reliability for Quantity Discrimination

Measure	Forms	*N*	*r*	*p*
Quantity Discrimination	QD 1 and QD 2	13	.84	.00
	QD 2 and QD 3	13	.74	.00
	QD 3 and QD 4	13	.58	.04
	QD 4 and QD 5	13	.80	.00
	QD 5 and QD 6	13	.78	.00
	QD 6 and QD 7	12	.88	.00
	QD 7 and QD 8	10	.74	.02
	QD 8 and QD9	11	.80	.00
	QD 9 and QD 10	10	.81	.01
	QD 10 and QD 11	9	.71	.03

Table 5

Alternate-form Reliability for Mixed Numeracy

Measure	Forms	*N*	*r*	*p*
Mixed Numeracy				
	MX 1 and MX2	13	.66	.01
	MX 2 and MX 3	12	.32	.32
	MX 3 and MX 4	11	.43	.19
	MX 4 and MX 5	12	.63	.03
	MX 5 and MX 6	13	.72	.01
	MX 6 and MX 7	13	.79	.00
	MX 7 and MX 8	13	.85	.00
	MX 8 and MX9	13	.80	.00
	MX 9 and MX 10	13	.70	.01
	MX 10 and MX 11	8	.62	.10

Examining the alternate-form reliability correlations for the Mixed Numeracy progress monitoring assessments, we found seven of the ten coefficients were statistically significant, ranging from .63 to .85.

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

The weekly growth rate data for the two progress monitoring measures are displayed in Table 6. Growth rates on the Quantity Discrimination assessments ranged from .01 to 1.03 correct responses per week, with a mean rate of improvement of .54 correct responses per week.

Mixed Numeracy growth rates ranged from .10 to 1.48 correct responses per week, with a mean rate of improvement of .66 correct responses per week.

Table 6

Weekly Growth Rates

		Slopes	
Student	Teacher	Quantity Discrimination	Mixed Numeracy
1	1	.28	.10
2	1	.78	.73
3	1	.90	.49
4	2	.65	1.48
5	2	.60	.85
6	3	1.03	.88
7	3	.54	.48
8	3	.17	.65
9	3	.89	.41
10	4	.50	.72
11	4	.23	.56
12	4	.50	.75
13	4	.01	.46

We used the Wilcoxon Signed Ranks Test to determine if student performance improved significantly from the beginning to the end of the progress monitoring period. For each of the measures we used the mean of the first two scores and the mean of the last two scores for these

comparisons. Students' scores improved significantly for both sets of measures (Quantity Discrimination [$z = -3.11$, $p < .01$]; Mixed Numeracy [$z = -3.19$, $p < .01$]).

Method: Intervention

In addition to determining how well the Quantity Discrimination and Mixed Numeracy 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 3-Tier Mathematics Model (3TMM) materials for the intervention sessions 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 2 and Teacher 3 participated in the TD intervention, while the students in the classes taught by Teacher 1 and Teacher 4 participated in the 3TMM intervention. Both groups of students were monitored using the progress monitoring measures described earlier in this report. All teachers used the same intervention that they had used in the preceding year (see Lind, Foegen, & Olson, 2009).

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 six students in the TD intervention group and seven 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 7 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 *U* Test. The results revealed that the groups were not significantly different on three of the measures (Number Identification, Quantity Discrimination, and Mixed Numeracy). Students in the TD group were significantly lower on the Missing Number task than students in the 3TMM group.

Table 7

Comparison of Fall Benchmarking Scores

	TD			3TMM				
Fall Screening	*n*	*M*	*SD*	*n*	*M*	*SD*	*U*	*p*
Number Identification	6	14.08	3.84	7	19.21	7.91	12.50	.23
Quantity Discrimination	6	12.92	5.70	7	17.50	2.61	8.00	.07
Missing Number	6	6.42	2.84	7	11.00	2.29	3.50	.01
Mixed Numeracy	6	13.42	3.76	7	17.21	1.60	8.50	.07

Next, we examined the growth exhibited by the two groups of students on the two progress monitoring measures to see if the changes in scores from the beginning to the end f the progress monitoring period were statistically 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 progress monitoring), but we split the file by group for this analysis. The results, presented in Table 8, show that students in both intervention groups demonstrated significant improvement from the beginning to the end of the progress monitoring period on both measures. We 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 ($U = 13.00$, $p = .30$) and for Mixed Numeracy ($U = 14.00$, $p = .37$).

Table 8

Comparison of Beginning and Ending Progress Monitoring Scores

		TD					3TMM			
Measure	*n*	Pre *M*	Post *M*	*z*	*p*	*n*	Pre *M*	Post *M*	*z*	*p*
Quantity Discrimination	6	17.58	36.33	-2.20	.03	7	28.79	35.64	-2.20	.03
Missing Number	6	18.17	31.42	-2.21	.03	7	22.71	32.57	-2.37	.02

Finally, we wanted to see if the one of the interventions led to higher scores on the spring benchmarking measures than the other intervention. The results presented in Table 9 reveal no significant differences between the spring benchmarking scores for the two groups.

Table 9

Comparison of Spring Benchmarking Scores

	TD			3TMM				
Spring Screening	*n*	*M*	*SD*	*n*	*M*	*SD*	*U*	*p*
Number Identification	6	40.25	10.92	7	45.21	11.01	16.00	.53
Quantity Discrimination	6	32.33	8.47	7	32.64	8.00	20.00	.95
Missing Number	6	16.00	4.06	7	17.64	3.41	16.00	.53
Mixed Numeracy	6	29.42	5.58	7	32.36	5.37	14.00	.37

Discussion

We were encouraged by the alternate-form reliability correlations obtained when the Quantity Discrimination measure was used for progress monitoring. Unlike the previous academic year when only two alternate-form reliability coefficients were close to or above the conventional .80 level (Lind, Foegen, & Olson, 2009), a majority of the coefficients for the present study met this standard. Although these were not quite as high as the alternate-form reliability for Quantity Discrimination when it was used as a benchmarking measure (Olson, Foegen, & Singamaneni, 2009), we were encouraged by the results obtained in this study.

This was the first year we used the Mixed Numeracy measures for both benchmarking and progress monitoring. The alternate-form reliability coefficients for benchmarking were all at or above the .80 level; as noted in Table 5, the progress monitoring coefficients were not as high. In general the coefficients improved as students and teachers became more familiar with the measure.

During this study, both of the ENIs used for progress monitoring were administered throughout most of the academic year. Both of the measures appear to be sensitive to growth, as the growth rates for the Quantity Discrimination and Mixed Numeracy measures were strong even though the data were drawn from a sample of first grade students who were struggling with mathematics concepts (and therefore represented a restricted range of ability).

Improvements in student scores resulted from both intervention approaches. As reported earlier, there were no significant differences between each group's rates of growth or spring screening scores. Even though there were no significant differences between the TD and 3TMM intervention, the teachers using 3TMM materials did find the materials easier to use than they had been the previous year; they would recommend them to others in the future.

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.

Impecoven-Lind, L., Olson, J., & Foegen, A. (2009). *Exploring the use of early numeracy indicators for monitoring progress in two intervention contexts* (Technical Report 28**).** 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.

Ngarkar, S., Hampton, D. D., Lembke, E. S., & Whitaker, T. (2009). *Missouri Early Numeracy Indicator screening data: 2007-2008.* (Technical Report XX). Minneapolis, MN: University of Minnesota, College of Education and Human Development, Research Institute on Progress Monitoring.

Olson, J., Foegen, A., & Singamaneni, S. (2009). *Iowa Early Numeracy Indicator screening data: 2007-2008* (Technical Report 23). Minneapolis, MN: University of Minnesota, College of Education and Human Development, Research Institute on Progress Monitoring.

Appendix A

Early Numeracy Indicators

Quantity Discrimination

Sample Quantity Discrimination Measure Page

Mixed Numeracy

Sample Mixed Numeracy 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

Mixed Numeracy, page 1 - Student Copy

4	7	2	1
12 | 9	6 | 1	3 | 8	10 | 7
2 3 ___ 5	4 5 6 ___	1 ___ 3 4	7 8 9 ___
41	8	21	11
18 | 9	6 | 10	20 | 15	1 | 7
5 10 15 ___	___ 5 6 7	3 4 ___ 6	20 30 40 ___
14	81	21	50

RIPM Research Institute on Progress Monitoring OSEP Award #H324H030003