School Psychology Review,
2009, Volume 38, No. 3, pp. 402–418

Sensitivity to Growth Over Time of the Preschool Numeracy Indicators With a Sample of Preschoolers in Head Start

Robin L. Hojnoski
Lehigh University

Benjamin Silberglitt
Technology and Information Educational Services, St. Paul, Minnesota

Randy G. Floyd
The University of Memphis

Abstract. There has been increased attention to the development of measures for assessing mathematical skill and knowledge in young children. Most of the evidence supporting these measures is consistent with Stage 1 research in the development of progress monitoring measures (Fuchs, 2004) and consists of investigation of technical features of performance at one point in time. The purpose of the current study was to move into Stage 2 research and examine sensitivity to growth over time of the Preschool Numeracy Indicators (PNIs; Floyd, Hojnoski, & Key, 2006) in a sample of Head Start preschoolers through a longitudinal design. Results indicated the PNI Oral Counting Fluency, One-to-One Correspondence Counting Fluency, Number Naming Fluency, and Quantity Comparison Fluency task scores are sensitive to growth over time and provide preliminary support for the promise of such measures in assessing early mathematical skill development. Consideration is given to implications for assessing early mathematical skill development in the context of general outcome measurement.

There has been increased attention to the development of measures for assessing mathematical skill in young children (e.g., Chard et al., 2005; Clarke, Baker, Smolkowski, &

A faculty research grant from The University of Memphis to the first author provided financial support for study. The opinions expressed in this article do not necessarily reflect those of the University of Memphis. We are grateful to Jessica Hall, who coordinated data collection and data entry efforts and participated in the assessment; to Ashley Smith, Missy Flinn, James Ford, Lauren McDurmon, and the other students who assisted with data collection; to the Head Start administrators and teachers who welcomed us into their center; and to the children who participated in data collection.

Correspondence regarding this article should be addressed to Robin Hojnoski, Department of Education and Human Services, Lehigh University, 111 Research Drive, Bethlehem, PA 18015-4794; E-mail: roh206@lehigh.edu

Chard, 2008; Floyd, Hojnoski, & Key, 2006; Methe, Hintze, & Floyd, 2008; Reid, Morgan, DiPerna, & Lei, 2006; VanDerHeyden, Broussard, & Cooley, 2006). Evidence is building to warrant such attention (e.g., National Mathematics Advisory Panel, 2008; National Center for Education Statistics, 2007) as mathematical skill appears critical to later school success. A recent meta-analysis of six longitudinal data sets suggests that mathematical skill at kindergarten entry is a strong predictor of later academic achievement, stronger than reading skills, attentional skills, or social behavior at kindergarten entry (Duncan et al., 2007). This suggests early mathematical competency is a primary target in improving academic achievement, and thus an important area for early identification and intervention.

Effective assessment practices play a critical role in promoting early identification and intervention in the development of mathematical competency. Assessment tools specifically designed for young children and targeting elements thought to provide a foundation for later success can provide data about acquisition of key skills as well as growth over time to inform effective instruction and intervention. In response to the limitations of more traditional early childhood assessment practices, curriculum-based assessment approaches are increasingly being applied to the assessment of growth and development in very young children with significant potential for improving outcomes (Bagnato, 2005; McConnell, 2000; VanDerHeyden, 2005; VanDerHeyden & Snyder, 2006).

General Outcome Measurement

Curriculum-based assessment can be conceptualized as the umbrella term for a number of assessment approaches characterized by key features of authenticity, instructional and intervention utility, reliability and validity, sensitivity to growth over time, and decision-making utility (Hintze, 2008). Within the larger domain of curriculum-based assessment, assessment approaches can be grouped as either specific subskill mastery measurement or general outcome measurement (GOM). Although both specific subskill mastery measurement and GOM approaches reflect the key features of curriculum-based assessment, there is a distinct difference in the development and content of each type of assessment. Whereas subskill mastery assessments typically and comprehensively sample a domain or hierarchy of skills that reflect an instructional sequence, in GOM a limited number of key skills are selected for measurement from the universe of possible skills and used as an indicator of global performance (Fuchs & Deno, 1991; Greenwood, Walker, Carter, & Higgins, 2006). In curriculum-based measurement (CBM; Deno, 1985, 1986; Shinn, 1989), perhaps the most widely known example of GOM, the focus is on broad, long-term objectives as opposed to the short-term objectives characteristic of a subskill mastery approach (Hintze, 2008). Moreover, the emphasis in CBM is on fluency with the target skill or behavior as measured across the assessment period, as opposed to mastery of specific skills in a hierarchy. CBM tasks, or tasks developed in the GOM framework, are designed as indicators of growth toward a long-term objective; thus, they are sensitive to growth over time and to the effects of instruction and intervention (Fuchs & Deno, 1991; Hintze, 2008; McConnell, Priest, Davis, & McEvoy, 2002). These features make GOM well suited for progress monitoring.

Development and Evaluation of Progress Monitoring Tools

To demonstrate the utility and viability of GOM for progress monitoring, three research stages must be completed (Fuchs, 2004). The first stage consists of investigating the technical features of performance at one point in time. This stage may include examining interscorer, alternate-form, internal consistency, and test–retest reliability as well as concurrent and predictive validity. Establishing the technical features of scores from one point in time is critical in ensuring that meaning can be attributed to the scores. Stage 2 research consists of investigating the technical features of slope, which may include the sensitivity of the measures to growth over time and the relation between growth on the measures and improvement in the target domain overall. An essential feature of GOM is the degree to which it is sensitive to performance changes that result from skill acquisition. Stage 3 research examines the instructional utility of the measures to determine whether the measures can be used to inform instruction and intervention. Both Stage 2 and Stage 3 research are particularly critical in considering the use of progress monitoring measures in tiered models of assessment and intervention (e.g., response to intervention).

Measuring Early Mathematical Development: Stage 1

Several research teams have begun to examine the use of GOM within a framework for formative assessment in early mathematical development. Creating a GOM for mathematics is challenging given the multiple strands that comprise mathematics (e.g., numbers and operations, algebra, geometry, measurement, data analysis). In early mathematics, the range of assessment targets is complicated further by the learning trajectories through which children progress and the variability in young children's development of mathematics (Clements, 2004). For example, in counting, roughly between the ages of 2 and 5, young children progress from nonverbally representing a collection of objects to counting aloud in sequence from a number other than 1 (Clements, 2004). Further, because young children have not developed the computational skills that are typically the focus of progress monitoring measures in elementary school, measures developed for young children have focused primarily on components of *number sense*.

There are several definitions of *number sense*, including "fluidity and flexibility with numbers, the sense of what numbers mean" (Gersten & Chard, 1999, p. 19). The National Council for Teachers of Mathematics standards for prekindergarten to Grade 2 suggest that a child's sense of number includes "understandings of size of numbers, number relationships, patterns, operations, and place value" (National Council for Teachers of Mathematics, 2000, p. 78). Independent of the precise definition, there is general agreement that number sense includes skills such as counting, making numerical comparisons, verbal and nonverbal calculations, estimations, and facility with number patterns (Berch, 2005); researchers investigating the use of GOM in early mathematical development have typically included tasks that reflect a subset of these skills.

Recently, 12 studies have been published detailing evidence for GOM tasks for early mathematics with samples of kindergarten and preschool students. Seven studies present a variety of technical adequacy evidence consistent with Stage 1 research that support the use and interpretation of measures of early numeracy, whereas 4 studies also present evidence of Stage 2 research. One study presents preliminary evidence of sensitivity to intervention with preschoolers (VanDerHeyden et al., 2006). Two studies by Clarke and colleagues (Chard et al., 2005; Clarke & Shinn, 2004) examined the interrater, alternate-form, and test–retest reliability as well as predictive validity of four early numeracy CBM (EN-CBM) measures for kindergarten and first-grade students. In the first study of the EN-CBM tasks (Clarke & Shinn, 2004), although reliability coefficients varied, all were strong for each type of reliability investigated. In addition, median concurrent validity correlations ranged from .60 to .75, and predictive validity coefficients were moderate to strong. Results from Chard et al. (2005) replicated the findings of Clarke and Shinn (2004) with regard to predictive and concurrent validity and subsequent studies (Lembke & Foegen, 2009; Lemke, Foegen, Whittaker, & Hampton, 2008; Martinez et al., 2009) have supported the strong technical properties of the EN-CBM.

In another study with kindergarten students, Methe et al. (2008) investigated the test–retest reliability, internal consistency, concurrent validity, and decision accuracy of four tasks measuring numerical and non-numerical knowledge. Two of the measures, Ordinal Position Fluency and Number Recognition Fluency, demonstrated reliability, validity, and cutoff score data appropriate for accurate classification decisions. Counting-on Fluency and Match Quantity Fluency, although relatively accurate in grouping decisions, had lower reliability and validity coefficients.

In a series of studies with preschool and kindergarten students, VanDerHeyden and colleagues examined alternate-form reliability, criterion-related validity, convergent and discriminant validity, decision-making accuracy, and predictive validity of a variety of early numeracy measures (VanDerHeyden et al., 2006; VanDerHeyden, Witt, Naquin, & Noell, 2001; VanDerHeyden et al., 2004). In general, results from the studies indicated strengths in alternate-form reliability and correlations with teacher ratings. Mixed results were found for other evidence of validity. For example, for the preschool tasks, correlations with the criterion measures ranged from low to moderate, and for the kindergarten tasks, correlations with criterion measures ranged from low to moderate and were not domain specific. Discriminant functional analyses indicated kindergarten measures were predictive of retention, problem validation procedures, and referral to a team.

Reid et al. (2006) developed a set of tasks targeting early literacy and early numeracy for use with a Head Start sample, and the internal consistency, item difficulty, and item discrimination of the probe scores were investigated. According to the authors, for the mathematics tasks, adequate item and scale properties were demonstrated and item-discrimination indices were positive and high. Adequate alternate-form reliability was also demonstrated and concurrent validity with the criterion measure was moderate to strong.

Finally, Floyd et al. (2006) developed four tasks as indicators of mathematical skill for preschoolers. Technical features of the measures were examined, including test–retest reliability, and four types of validity evidence—content, response processes, internal relations, and external relations—were examined with samples of children from diverse preschool settings. Evidence of the technical properties discussed under Measures indicated sufficient support to continue to explore additional features of the measures consistent with Stage 2, such as growth over time.

At least two patterns are evident across this growing body of research. First, in terms of content, although the domain of mathematics consists of various strands as reflected in the National Council for Teachers of Mathematics standards (e.g., numbers and operations, geometry, algebra, measurement, and data analysis and probability), tasks primarily focus on numbers and operations, with few exceptions. This results in a great deal of similarity among tasks across research teams. For example, each of the research teams used some variation of a number identification task and a counting task. Second, in terms of supporting evidence for the measures, consistent with the larger body of research on mathematics CBM tasks (Foegen, Jiban, & Deno, 2007), much of the research on early mathematics GOM tasks has focused on Stage 1 research, with some variability in the technical features investigated. In general, this collective body of research suggests that it is possible to develop technically adequate measures of early math development. A logical and necessary next step is to examine features of developed measures that are consistent with Stage 2 research. Initial efforts have been made in this direction.

Measuring Early Mathematical Development: Stage 2

Four studies examine the EN-CBM tasks (Clarke & Shinn, 2004), or a modified version of these tasks, and provide evidence for GOM tasks in early mathematics consistent with Stage 2 research. Clarke et al. (2008) examined the technical features of slope, including the predictive validity of slope over time and its unique contribution to predicting performance when examined with other measures. EN-CBM (Clarke & Shinn, 2004) and criterion measures were administered at the beginning and end of the year to a sample of 254 kindergarteners. EN-CBM tasks (Clarke & Shinn, 2004) were administered at three additional points during the year to a subsample of kindergarteners to compare the predictive power of slope to the static score from the criterion measures. Results indicated that only one measure, Quantity Discrimination, fit a linear growth curve, and similarly, growth on only Quantity Discrimination accounted for variance above and beyond static criterion measures. In addition, Lembke and colleagues investigated three of the Clarke and Shinn (2004) tasks as well as modified versions of the tasks. Lembke et al. (2008) reported that Quantity Discrimination, Missing Number, and Number Identification demonstrated sensitivity to growth when administered to 77 kindergarten and 30 first-grade students at approximately 4-week intervals across seven rounds of data collection. Both linear and polynomial models revealed that children's performance improved across time on the three measures. Results indicated significant linear growth for only Number Identification. Growth in Quantity Discrimination and Missing Number was significant but nonlinear. Lembke and Foegen (2009) examined modified versions of the tasks in a sample of 72 kindergarten and first-grade students. Differences between fall and spring scores were calculated to determine mean growth across the time periods, and were tested using paired sample *t* tests. Although both kindergarten and first-grade students improved over the course of time, mean differences were statistically significant for kindergarten students only. In a study with only kindergarten students using three of the EN-CBM tasks (Clarke & Shinn, 2004), students improved on all three tasks from fall to spring. Further, repeated-measures analysis of variance indicated a significant and large effect for time (Martinez et al., 2009).

The results of these Stage 2 investigations indicate some of the early numeracy tasks show promise, although results are not consistent across all tasks for all samples, and the degree to which GOM tasks can be used for progress monitoring remains unclear. It is important to note that all of the Stage 2 investigations were conducted with kindergarten and first-grade students, and all of them investigated EN-CBM, or a modified version of the measures. Although these studies provide estimates of growth over time to support the use of the EN-CBM, research is needed to understand whether sensitivity to growth is a characteristic of the early mathematics GOM tasks that have been developed for preschoolers. Investigating sensitivity to growth over time in early mathematics GOM tasks for preschoolers is particularly important given that differences in mathematical competencies appear by the age of 3 (Case, Griffin, & Kelly, 1999) and that mathematical competency at kindergarten entry is a strong predictor of later achievement (Duncan et al., 2007).

Purpose of the Study

The purpose of the present study was to advance Stage 2 research on early mathematics GOM tasks specifically with preschoolers by examining the degree to which the tasks created by Floyd et al. (2006) demonstrate sensitivity to growth over time via a longitudinal study that purposefully included a range of ages. Measuring the same students over time provides more definitive evidence regarding sensitivity to growth and represents the next logical step in Stage 2 research on the development of a GOM for early numeracy. In addition, the present study focused on the ability of the tasks to detect changes in performance over time in a sample of preschoolers from low-income ethnic minority families attending a Head Start program. Efforts to create measurement tools that can be used successfully with populations in which performance disparities are evident (National Center for Education Statistics, 2007; National Mathematics Advisory Panel, 2008) are critical to improving mathematical outcomes for young children. The general goal of the study was to examine how the tasks developed by Floyd et al. (2006) functioned as early mathematics GOM tasks. These measures were selected as the focus for two reasons. First, the PNIs demonstrated sufficient technical properties to support continued examination consistent with Stage 2 research. Second, several of the tasks are conceptually aligned with EN-CBM (Clarke & Shinn, 2004). That is, several of the

PNI tasks are similar in content to EN-CBM in a form modified to be appropriate for preschoolers. Demonstrating growth over time of the PNIs is an essential first step in establishing a continuous measurement system that links with EN-CBM. The specific research questions of the current study were as follows: (a) What are the basic descriptive features of each task (e.g., mean, standard deviation, skewness, kurtosis, and percent of zero scores)? (b) What is the degree to which each task reflects growth over a measurement period of 8 months, or approximately one school year? (c) What is the degree to which each task is sensitive to potential interactions with child age and classroom placement?

Method

Participants and Settings

Participants were 139 children enrolled in an urban Head Start program located in the mid-South. From this sample, 66 participants were boys and 73 participants were girls. At the beginning of the study, the children ranged in age from 37 months to 60 months (M = 51.0, SD = 6.1 month). All of the children were African American, and English was their primary language. None of the participants had been diagnosed formally with a disability condition.

The Head Start program was located in a private agency that also operated a day care center and a state-funded preschool program, although only children attending the Head Start program participated in the study. The Head Start program consisted of seven classrooms, each serving approximately 20 children and each staffed with one teacher and one teacher assistant. Children were assigned to one of seven classrooms by the family services manager; no information was provided about the specific procedures used to assign children. Because assignment to classrooms was not specified, classrooms included unequal numbers of boys and girls as well as unequal numbers of 3-, 4-, and 5-year-olds.

According to the education manager, the Head Start program used the Creative Curriculum for Preschool, Fourth Edition (Trister Dodge, Colker, & Heroman, 2002), which has curriculum objectives aligned with the National Council for Teachers of Mathematics standards in number concepts, patterns and relationships, geometry and spatial sense, measurement and data collection, organization, and representation. The objectives are broad and many are the same across the content of the standards. For example, one objective indicates the child will observe objects and events with curiosity whereas another indicates the child will approach problems flexibly. In addition, all of the classrooms followed Head Start performance standards in designing classroom activities and instruction. Standards were reviewed regularly at staff meetings to ensure teachers were implementing instruction consistent with the standards across specific domains. Teachers followed similar general classroom schedules and informal observations suggested teachers used many of the same types of learning activities (i.e., circle time, learning centers); however, there was some variability in implementation. For example, some teachers made different learning centers available at different times, and different materials were evident in each of the classrooms. In terms of assessment, classroom teachers assessed child progress at three points in time during the school year using the Learning Accomplishment Profile—Third Edition (Sanford, Zelman, Hardin, & Peisner-Feinberg, 2003).

Measures

Four PNIs (Floyd et al., 2006) were administered to all children to assess early mathematical development.[1] All PNIs but Oral Counting Fluency include demonstration and sample items designed specifically to minimize scores of 0 that are not reflective of true skill levels.

One-to-One Correspondence Counting Fluency targets the ability to count objects fluently and requires children to point to and count circles approximately 1-inch in diameter printed on a page. Circles are presented in four rows of 5 circles. Children have a maximum of 30 s to count all 20 circles. The Fluency score results from multiplying 30 times the number associated with the last circle counted only once in sequence and dividing the product by the child's time of completion. Time of completion for all children who do not correctly count all circles is 30 s. Prior research revealed test–retest reliability coefficients of .62 across a 2- to 4-week interval and .96 across a 5- to 7-week interval. Corrected correlations between One-to-One Correspondence Counting Fluency and scores from the Bracken Basic Concept Scale—Revised (Bracken, 1998) and the Woodcock-Johnson III Applied Problems test (Woodcock, McGrew, & Mather, 2001) ranged from .29 to .38. Corrected correlation with the Test of Early Mathematics Ability—Third Edition (Ginsburg & Baroody, 2003) total scores was .64 (Floyd et al., 2006). A 1-month test–retest reliability coefficient from the current sample (from the first assessment to the second) was .77.

Oral Counting Fluency targets the ability to produce numbers fluently in sequence, beginning with the number 1. Children are asked to state numbers in sequence until they reach the highest number they can produce in 1 min. The Fluency score represents the number of numbers stated correctly in sequence from 1 in 1 min. Prior research revealed test–retest reliability coefficients of .90 across a 2- to 4-week interval and .82 across a 5- to 7-week interval. Corrected correlations between Oral Counting Fluency and scores from the Bracken Basic Concept Scale—Revised and the Woodcock-Johnson III Applied Problems test ranged from .31 to 45. Its corrected correlation with Test of Early Mathematics Ability—Third Edition total scores was .55 (Floyd et al., 2006). The 1-month test–retest reliability coefficient from the current sample was .71.

Number Naming Fluency targets the ability to name numerals fluently. It requires children to say the names of the numerals 0–20 presented one at a time. Examiners present these pages in rapid succession for 1 min, and children have 3 s to respond to each page. The 21 numerals included in the task are presented in random sequences in three complete sets for a total of 63 possible presentations. The Fluency score represents the number of numerals named in 1 min. Prior research revealed test–retest reliability coefficients of .91 across a 2- to 4-week interval and .88 across a 5- to 7-week interval. Corrected correlations between Number Naming Fluency and scores from the Bracken Basic Concept Scale—Revised and the Woodcock-Johnson III Applied Problems test ranged from .29 to .40. Its corrected correlation with the Test of Early Mathematics Ability—Third Edition total scores was .70 (Floyd et al., 2006). The alternate-form reliability coefficient at a 1-month interval from the current sample was .72.

Quantity Comparison Fluency targets the ability to make judgments fluently about differences in the quantity of object groups. It requires children to identify which of two boxes of circles printed on a page contains more circles. Children respond by touching the box with more circles. Each side contains 1–6 circles, and each quantity of circles is represented in a standard fashion across pages. Examiners present, in rapid succession, up to 30 pages with sets of circles on each side for 1 min. Children have 3 s to respond to each page. The Fluency score results from multiplying 60 times the number of correct responses and dividing the product by the child's time of completion (1 min or less). Prior research revealed test–retest reliability coefficients of .89 across a 2- to 4-week interval and .94 across a 5- to 7-week interval. Corrected correlations between Quantity Comparison Fluency and scores from the Bracken Basic Concept Scale—Revised and the Woodcock-Johnson III Applied Problems test ranged from .38 to .58. Its corrected correlation with the Test of Early Mathematics Ability—Third Edition total scores was .58 (Floyd et al., 2006). The alternate-form reliability coefficient at a 1-month interval from the current sample was .65.

Procedure

Training and reliability. Three undergraduate students in psychology and seven graduate students in school psychology who were trained to criterion in administering all measures collected the data. Initial training consisted of two 1-hr group sessions, one individual session, and a practice administration. Monthly group meetings were held to discuss any difficulties, questions, or concerns that arose during administration and to minimize assessor drift. In addition, the first author performed at least one assessment integrity check with each examiner during the data collection process using a procedural checklist developed by the first author. All administrations observed were completed with 100% integrity.

Data collection. The PNIs were administered as part of a larger assessment that included the Individual Growth and Development Indicators (McConnell et al., 2002) and a letter-naming task. Only results pertaining to the PNIs are reported here. All tasks were individually administered by the trained examiners and completed in the child's classroom at a child-sized table apart from other classroom activities. Assessment batteries were counterbalanced across children to minimize order effects in performance. The entire administration was completed in approximately 10 min. Children received a sticker when the assessments were completed.

Data collection was part of a larger effort initiated by the Head Start program to monitor the children's development in the key preacademic areas of early literacy and early numeracy. All children who attended the Head Start center participated in the data collection, and caregivers were notified of the process by the Head Start program. No parents objected to their children's participation and thus no children were excluded from the study unless they left the program while data were being collected.

Data were collected monthly on all children who were in attendance on data collection sessions from October to May. The assessments were scheduled at equal intervals across the entire data collection period. A monthly schedule of assessments was selected to determine a precise rate of growth for short time intervals. In a typical benchmark assessment strategy, three periods of assessment are used to track student performance over time. However, although benchmark assessment strategies are quite common in the kindergarten through 12th-grade setting, there is no corresponding standard for GOM assessment in a preschool setting. Monthly assessments provided an appropriate balance between the frequency of assessment that might typically happen in a preschool setting and collecting a sufficient number of data points to limit the standard error of the slope estimate (*SE*b).

Data Analysis Plan

One goal of this study was to examine how the PNIs functioned as a progress monitoring tool according to basic descriptive features of the tasks as well as growth over time. First, descriptive statistics were examined for the potential early mathematics GOM tasks used in the study. Previous research on children in the preschool-age range has found a significant number of zero scores for the Individual Growth and Development Indicators, indicating a floor effect for these measures (Missall et al., 2007). It was important, then, to determine whether a floor effect might also be present for the PNIs with this sample.

Second, to explore the growth of each task, sensitivity to growth and potential interactions with child age and classroom placement were examined using a linear mixed model (LMM). It is important to use an analysis tool that allows for explicit modeling of individual differences in growth (Raudenbush & Bryk, 2002). LMM, of which hierarchical linear modeling is a well-known example, are especially suited to the type of problem addressed by this study (Fitzmaurice, Laird, & Ware, 2004). LMMs are useful for analyzing longitudinal data, where the assumption of independence between measurements cannot be met. Measurements are nested within children, and in this case, children are nested within their age and classroom placement. The random effects structure allows individual children to vary on all parameters being estimated while still providing group-level estimates that are more appropriate than those generated by more traditional regression models.

In this analysis, the Level 1 equations model individual growth over time. For each individual i at time j, the Level 1 linear equation is as follows:

$$y_{ij} = \beta_{0i} + \beta_{1i}t_{ij} + e_{ij} \quad (1)$$

In Equation 1, linear growth is modeled for each student, with e_{ij} representing error. The Level 1 quadratic equation is as follows:

$$y_{ij} = \beta_{0i} + \beta_{1i}t_{ij} + \beta_{2i}t_{ij}^{2} + e_{ij} \quad (2)$$

The Level 2 equations then model group differences in key beta values. For the purposes of explanation, Level 2 equations will be presented that model group differences in β_{1i}, the linear slope estimate for individual i in Equation 1. Level 2 equations could also be examined for all beta values in any of the Level 1 equations. Equations 3 and 4 represent Level 2 equations for differences in classroom group mean estimates (similar equations were used for differences in age). In these equations, c is a dummy-coded variable reflecting whether the child participated in that classroom, and b is the error term in each equation.

$$\beta_{0i} = \beta_0 + \beta_2 c_{1i} + \beta_3 c_{2i} + \ldots + \beta_n c_n i + b_0 i \quad (3)$$

$$\beta_{1i} = \beta_1 + \beta_{10} c_{1i} + \beta_{11} c_{2i} + \ldots + \beta_n c_n i + b_{1i} \quad (4)$$

LMMs also allow for missing data, such as those from attrition. Missing data are a common occurrence when conducting longitudinal studies. The percentage of missing data in this study ranged from 0.48% to 0.98% for each measure. Data were assumed to be missing at random. Similar approaches to handling missing data have been used in other studies investigating early mathematics performance over time (e.g., Jordan, Kaplan, Olah, & Locuniak, 2006). LMMs also assume independence between subjects, an assumption that can be addressed by sampling randomly from the population. Although this sample is from a single region of the country, we believe this assumption has been met for the sake of these analyses and recognize the limitations of our sample. A final assumption of LMMs is that the response variable be normally distributed. A more thorough discussion of the distributional properties of each measure is addressed later in this article. In those cases where the measure varied significantly from a normal distribution, analyses were still conducted, keeping in mind the limitations of the data set.

Results

Results are presented first in terms of the distributional properties of each measure. Next, the measures are examined for sensitivity to growth. Last, group-level differences for classroom and age are described.

Descriptive Statistics

The distributional properties of each measure were first explored. High standard deviations relative to the mean, large positive skewness values, large positive kurtosis values, and a high number of zero scores all raise concerns about a measure's ability to discriminate well across children. In this particular sample, some of these concerns are present with the PNIs. All data, regardless of wave of data collection, were compiled and both percentage of zero scores and descriptive statistics were explored. Students who did not successfully complete the sample items (i.e., error or no response), and students who successfully completed the sample items but did not correctly complete a single item, both received a zero score in this study. These data are presented in Table 1.

Next, students were divided as to whether they had reached 4 years of age at the time of measurement, and the percentage of zero scores across these two groups were compared. This analysis was to better understand whether the percentage of zero scores decreased when measuring older versus younger children. Percentage of zero scores was moderately high for Number Naming Fluency

Table 1
Descriptive Statistics of the Preschool Numeracy Indicators

Statistic	OOCCF	OCF	NNF	OCF
Percentage of zero scores				
Total	3.8	1.6	37.8	8.1
4+	0.8	0.5	30.9	4.8
<4	14.5	5.5	62.0	19.8
Descriptive statistics—All students				
M	16.6	20.3	5.1	15.7
Mdn	12	15	2	15
SD	16.3	15.1	7.1	8.4
Skewness	1.5	1.7	2.0	−0.1
Kurtosis	1.3	5.1	4.2	−0.5
Descriptive statistics—No zero scores				
M	17.3	20.6	8.1	17.1
Mdn	13	15	6	16
SD	16.3	15.0	7.5	7.3
Skewness	1.5	1.7	1.6	0.2
Kurtosis	1.2	5.2	2.6	−0.5

Note. OOCCF = One-to-One Counting Correspondence Fluency; OCF = Oral Counting Fluency; NNF = Number Naming Fluency; QCF = Quantity Comparison Fluency; 4+ = children 4 years old and older; <4 = children under the age of 4 at time of measurement.

(37.8%), with large differences between older and younger groups of children (30.9 vs. 60.2, respectively). The percentage of zero scores for all other measures was below 10% for the overall group and below 20% for the younger group.

Descriptive statistics of the measures were first examined with zero scores included, as shown in Table 1. These values revealed high standard deviations relative to the mean for most measures, although for Quantity Comparison Fluency, standard deviations were around half of the mean. Quantity Comparison Fluency also demonstrated the skewness values closest to zero, although skewness was moderately positive (between 1.5 and 2.0) for all other PNI tasks. In addition, both Quantity Comparison Fluency and Number Naming Fluency demonstrated highly leptokurtotic distributions.

Because previous research on other preschool GOMs (i.e., Individual Growth and Development Indicators; Missall et al., 2007) has often removed the zero scores before conducting analyses, descriptive statistics were again examined with zero scores removed, also shown in Table 1. The key descriptive statistics of the PNIs remained remarkably stable when comparing before and after the removal of zero scores; the largest change in skewness and kurtosis was in Number Naming Fluency and the largest change in standard deviation was in Quantity Comparison Fluency.

Sensitivity to Growth

The purposes of further analyses in this study were to examine sensitivity to growth and whether group differences could be modeled across age and classroom placement. In a practical setting, these measures will be administered to all children, not just those children who already demonstrate scores greater than zero. Thus, characteristics such as sensitivity and group differences pertained to all students in the sample, regardless of whether they achieved a score higher than zero on a given measure. The decision was made to con-

Table 2
Parameter Estimates for Linear and Quadratic Growth Models for the Preschool Numeracy Indicators

Parameter	OOCF	OCF	NNF	OCF
Linear model				
Intercept	12.4	16.5	3.5	12.2
Linear	1.37	1.15	0.52	1.09
Linear *p* value	<.001*	<.001*	<.001*	<.001*
Quadratic model				
Intercept	13.0	17.3	3.3	11.3
Linear	0.9	0.3	0.7	2.0
Quadratic	0.07	0.13	−0.03	−0.14
Quadratic *p* value	.47	.08	.15	.004*

Note. OOCCF = One-to-One Counting Correspondence Fluency; OCF = Oral Counting Fluency; NNF = Number Naming Fluency; QCF = Quantity Comparison Fluency.
* Statistically significant with Bonferroni correction.

duct all further analyses on the data set with zero scores included.

The sensitivity to growth of all measures was examined using the LMM. First, the model explored whether the measures demonstrated a significant linear growth component. Next, the model examined whether a significant quadratic component was present. Parameter estimates and significance values are given for both the linear-only condition and the quadratic condition in Table 2. Significance values of the linear component must be explored in a linear-only condition, as significance tests of lower order parameter estimates are not appropriate in the presence of higher order parameter estimates (Fitzmaurice et al., 2004).

All measures demonstrated a significant linear growth component. Linear parameter estimates were around 1 item correct per wave of data collection for One-to-One Counting Correspondence Fluency, Oral Counting Fluency, and Quantity Comparison Fluency. Average growth was around 0.5 items correct per wave for Number Naming Fluency. In addition, a significant negative quadratic component was found for Quantity Comparison Fluency whereas a nonsignificant but large positive quadratic component was found for Oral Counting Fluency. Finding a significant linear growth component is strong, supporting evidence for the sensitivity of the measures to growth over time. It indicates that the rate of change of children's scores on these measures over time was nonzero and positive, and that this difference from zero was greater than what could simply be accounted for by the standard error of the slope estimate (*SE*b).

Classroom and Age Effects

An exploratory analysis was conducted to evaluate the effects for classroom and age separately using a LMM with a random effects covariance structure.

Classroom effects were examined using a series of omnibus tests across all classroom conditions, to examine whether any significant differences across classrooms might be found for each measure. Although no intervention was used and the curriculum used in each classroom ostensibly was the same, informal observations suggested instructional variations occurred across classrooms because of differences in teaching styles. However, these instructional variations were not controlled or documented, and no attempt was made to support an expectation that a particular classroom would outperform another; therefore, this element of the data analysis plan remains exploratory. In the case of significant omnibus tests, individual classroom comparisons could be conducted, but were not. The purpose of this analysis was only to establish whether classroom differences in slope estimates might reasonably be examined using this model, so as to inform future research. Classroom assignment was assumed to be random.

After controlling for family-wise error using a Bonferroni correction ($p = .05/4 = .0125$), significant omnibus tests for classroom differences were found for Quantity Comparison Fluency ($p < .001$). Omnibus tests were near significance for Oral Counting Fluency ($p = .02$). This initial evidence supports the need for further exploration into the capability of the PNI tasks to differentiate growth differences across classrooms and interventions.

To examine the effect of age, children's initial age was entered into the model, using age at October 1 of the year the study began. After controlling for family-wise error again using a Bonferroni correction (p = .05/4 = .0125), a significant effect for child age was not found for any measure, with the exception of Number Naming Fluency (p = .004). Thus, for all measures except Number Naming Fluency, slope is unlikely to be significantly related to the age of the student, for the population represented by this sample. For Number Naming Fluency, older children demonstrated significantly steeper slopes. This indicates that growth rates tended to be higher for older versus younger participants and that these differences were greater than that which might be explained by the standard error of the slope estimate (*SE*b). A significant effect for age on slope would indicate that the measure was less sensitive to growth over certain age ranges and may be less appropriate for certain portions of the age range of participants in the study.

Discussion

The goal of the study was to examine how the PNIs functioned as an early numeracy GOM in terms of sensitivity to growth over time with a sample of children from predominantly low socioeconomic backgrounds. In terms of general descriptive statistics, the PNIs demonstrated relatively low percentages of zero scores across three of the four tasks. The exception was the high percentage of zero scores for Number Naming. The lower percentage of zero scores for three of the PNI tasks is encouraging in that a high percentage of zero scores coupled with large standard deviations relative to the mean call into question the ability of the task to differentiate between children on the lower end of the distribution, a necessary characteristic for an adequate screening measure. It is important to note that although there is a rationale for selecting a sample that may be at risk for mathematical development, such a sample makes it difficult to evaluate the descriptive properties of the tasks in the context of the utility of the PNIs for the larger preschool population. Additional research with diverse samples is needed to determine if a low percentage of zero scores is consistent across samples and if the high percentage of zero scores for Number Naming Fluency is an artifact of the sample or a characteristic of the task.

Despite the low percentage of zero scores for Quantity Comparison Fluency, Oral Counting Fluency, and One-to-One Counting Correspondence Fluency, there remain some concerns. The large standard deviations relative to the mean, combined with moderate positive skewness for all but the Quantity Comparison Fluency, and the high leptokurtosis for the Oral Counting Fluency and Number Naming Fluency, are often indicative of a floor effect. In addition, large standard deviations raise the issue of measurement error and the extent to which a true score can be identified. Especially in light of the standard error of measurement, in such distributions it becomes difficult to make meaningful statements about differences across students whose scores are at the bottom of the range of performance.

However, the skewness values were moderate, not severe, and the demographics of the current sample raise the possibility that performance was more homogenous and potentially different from that of the general population. Distinct patterns of performance have been demonstrated in research with samples of kindergarten children similar in demographics (Jordan et al., 2006; Jordan, Kaplan, Locuniak, & Ramineni, 2007). In addition, children did demonstrate significant growth on the measures across months in the Head Start environment. This lends further support to the notion that these distributional properties may be a function of the sample, as exposure to a rich instructional environment led students to grow up and off of any "floor" in the distribution. Thus, further research should explore the distributional properties of these tasks with a larger and more representative sample to establish whether these issues are a function of the limited range of this particular sample.

Sensitivity to Growth Over Time

Results suggest each PNI task is sensitive to growth over time, with growth rates for three of the tasks (i.e., Quantity Comparison Fluency, Oral Counting Fluency, and One-to-One Counting Correspondence Fluency) calculated at 1 item per month. Number Naming Fluency demonstrated growth over time—but at a rate of 0.5 items per month. These growth rates are encouraging in that they are of sufficient magnitude to be meaningful for progress monitoring. That is, if the PNIs were used on a benchmarking schedule, one could expect a growth rate of between 2 and 4 items per benchmark period, which is a growth rate large enough to be visually detected when graphing data. Additional research with more diverse samples is needed to determine whether growth rates differ. Steeper growth rates would facilitate the use of the tasks as progress monitoring measures that can be used more frequently than the typical benchmark schedule. Further, research is also needed to determine how growth over time is related to important early mathematical outcomes, similar to the research conducted by Clarke et al. (2008).

Classroom and Age Effects

Demonstrated differences for Quantity Comparison Fluency and Oral Counting Fluency across classrooms might be considered an initial indication of sensitivity to intervention in that, assuming assignment of children to classrooms is mostly random, different group mean slope estimates indicate that children may be experiencing different instructional environments. This finding suggests that Quantity Comparison Fluency and Oral Counting Fluency may be useful tools in evaluating the effects of instruction and curriculum, although caution should be used in interpreting this finding. Differences in instruction and curriculum were not systematically documented in this study, only informally noted. Stage 3 research efforts are needed to experimentally explore this area and demonstrate the utility of each of the PNI tasks. Such efforts should be conducted with careful attention to assignment to classrooms and intervention implementation and integrity. Sensitivity to intervention is critical for formative assessment measures to be used in tiered models of service delivery. Response to intervention is increasingly becoming part of the dialogue in early education, and the need for tools that are sensitive to growth over time and the effects of intervention are critical (VanDerHeyden & Snyder, 2006).

Differences in slope across initial age were generally not found, with the exception of Number Naming Fluency. Children demonstrated growth over time on these tasks, and typically this growth was similar for both older and younger children. This fact, combined with the relatively low number of zero scores for these tasks, suggests Oral Counting Fluency, One-to-One Correspondence Counting Fluency, and Quantity Comparison Fluency can reasonably be given to assess performance over time across the entire age range represented by this sample. Further research should explore performance across the age range with a broader sample, perhaps in a cross-cohort longitudinal design using linear growth modeling, to understand more fully if there are specific age ranges where a given PNI demonstrates steeper growth and thus greater sensitivity to changes in early numeracy skill.

Limitations

There are a number of limitations of the current study, the existing evidence for the PNIs, and aspects of the tasks that need to be addressed through future research. First, the sample for this study was highly specific in terms of geographic location, educational setting, race, at risk, and socioeconomic status. The specific characteristics of the sample prevent estimation of "typical" performance; it is possible that patterns of performance as well as growth over time may differ in other samples. To determine whether the measures hold promise for a more general population, further research is needed to establish whether similar growth rates would be demonstrated with a more diverse sample representative of the general population. Further research is also needed to better understand the potential differences between children who do not successfully complete the sample items and children who complete the sample items but do not correctly respond to any items in the task, both of whom received a score of zero in this study.

Currently, the evidence to support the PNIs is limited in scope, although growing. The National Center on Response to Intervention (n.d.) outlines necessary criteria for screening and progress monitoring tools. To certify the PNIs as a screening and progress monitoring tool, the following needs to be established: (a) classification data, (b) predictive validity, (c) benchmarks, (d) improved teacher planning or student achievement as a result of using the PNIs, (e) predictive validity, (f) sensitivity to the effects of intervention, and (g) reliability of slope. This information will inform recommendations as to which PNI tasks are most useful in an early warning system designed to improve outcomes for young children.

Finally, the PNIs, like most of the early numeracy measures developed across research teams, are limited in their focus. The PNIs represent tasks thought to relate to number sense with no attention to other areas of mathematical development (e.g., algebra, geometry, measurement, data analysis, and probability). Understanding how number sense relates to other areas is critical in validating number sense tasks as GOM for mathematical competency. To be considered a GOM, growth in single-skill measurement tasks, like those for number sense, must correspond to global learning in the broader domain of interest (Fuchs, 2004). In addition, future research should consider other areas of mathematical development for potential GOM tasks.

Implications for Practice

The purpose of the present study was to further Stage 2 research on early mathematics GOM tasks by examining the sensitivity to growth over time of the PNI tasks in a sample of preschoolers who may be at risk for academic difficulties because of socioeconomic status and ethnicity. Although the results of the study have limited generalizability, there are important implications for early educators, school psychologists, and other educational personnel in developing and implementing systems-level approaches to promote positive academic and social outcomes for children. First, although additional research is necessary with more diverse samples, this study provides preliminary evidence that the PNIs may be a useful progress monitoring tool for preschool-age children. Results indicate all of the PNI measures demonstrate sensitivity to growth over time with a sample of preschoolers at risk for mathematics difficulties, with Oral Counting Fluency, One-to-One Counting Correspondence Fluency, and Quantity Discrimination Fluency demonstrating greater growth rates than Number Naming Fluency. Further, although confirmatory research is needed, the growth rates from this study suggest promise for using Quantity Comparison Fluency and Oral Counting Fluency to monitor progress more frequently to evaluate response to intervention and inform data-based decision making about children's educational programming.

Second, this study used a longitudinal design that purposefully included the full age range of preschoolers (i.e., 3- to 5-year-olds), and results provide support for the use of Oral Counting Fluency, One-to-One Correspondence Counting Fluency, and Quantity Comparison Fluency with the full age range, eliminating the need for specific tasks to be used only with children of a specific age. Use across the preschool-age range facilitates tracking children's growth and development in early mathematics at a critical time point, given research that suggests differences in children's growth trajectories in mathematics emerge by 3 years of age and increase through the preschool years (Case et al., 1999; National Mathematics Advisory Panel, 2008). More research is needed to explore age differences in Number Naming Fluency and its utility across the age range.

Finally, this study was conducted purposefully with a sample of low-income minority preschoolers. Although conducting research with a broader sample may provide more information about typical performance and growth over time on the PNI tasks, demonstrating sensitivity to growth over time with a population at risk for mathematical difficulties based on demographics is critical. These children may not exhibit the same growth rates as the broader population, and thus tasks that detect growth for children at risk for mathematical difficulties are needed to inform instruction and intervention to efforts. Differences in performance between children from low-income households and their peers from higher income households provide evidence of the need for attention to these disparities at the earliest point possible (Jordan et al., 2006; National Council for Education Statistics, 2007; National Mathematics Advisory Panel, 2008).

In general, a set of early mathematics GOM tasks will provide educators and parents with important information about a child's growth trajectory and a means to better determine whether a child is developing the skills needed for mathematical competency. This early warning system can allow for early, intensive intervention, to help children gain the skills necessary and a means for monitoring progress toward valued outcomes in mathematical development. It remains unclear which of the PNI measures is the most useful and demonstrates the strongest properties as a GOM for use in screening or progress monitoring. At this point in the development of the PNIs, evidence has been demonstrated for alternate-form reliability, test–rest reliability, some forms of validity (Floyd et al., 2006), and sensitivity to growth over time with promise for sensitivity to the effect of instruction. Collectively, this evidence supports continued exploration of the PNIs in a data-based decision-making system through careful practical applications in the context of research and measurement development.

Footnotes

[1]Following collection of the data used in Floyd et al. (2006), PNI materials were revised as follows: (1) some wording in the directions was simplified; (2) decision rules for discontinuing the administration after completion of sample items were added; (3) test records and directions for each probe were separated; (4) boxes were added to Quantity Comparison Fluency to eliminate the concept of "side" of the page. Eight alternate forms of Number Naming Fluency and Quantity Comparison Fluency were prepared for repeated measurement. (No alternate forms are needed for Oral Counting Fluency and One-to-One Correspondence Counting Fluency.) Technical properties are reported for the tasks used in Floyd et al. (2006), except where noted.

References

Bagnato, S. J. (2005). The authentic alternative for assessment in early intervention: An emerging evidence-based practice. *Journal of Early Intervention, 28,* 17–22.

Berch, D. B. (2005). Making sense of number sense: Implications for children with mathematical disabilities. *Journal of Learning Disabilities, 38,* 333–339.

Bracken, B. A. (1998). *Bracken Basic Concepts Scale—Revised.* San Antonio, TX: Psychological Corporation.

Case, R., Griffin, S., & Kelly, W. M. (1999). Socioeconomic gradients in mathematical ability and their responsiveness to intervention during early childhood. In D. P. Keating & C. Hertzman (Eds.), *Developmental health and the wealth of nations: Social, biological, and educational dynamics* (pp. 125–149). New York: Guilford Press.

Chard, D., 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,* 3–14.

Clarke, B., Baker, S., Smolkowski, K., & Chard, D. J. (2008). An analysis of early numeracy curriculum-based measurement: Examining the role of growth in student outcomes. *Remedial and Special Education, 29,* 46–57.

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.

Clements, D. H. (2004). Major themes and recommendations. In D. Clements, J. Sarama, & S. DiBiase (Eds.), *Engaging young children in mathematics: Standards for early childhood mathematics education* (pp. 7–72). Mahwah, NJ: Erlbaum.

Deno, S. L. (1985). Curriculum-based measurement: The emerging alternative. *Exceptional Children, 52,* 219–232.

Deno, S. L. (1986). Formative evaluation of individual student programs: A new role for school psychologists. *School Psychology Review, 15,* 358–374.

Duncan, G. J., Claessens, A., Huston, A. C., Pagani, L. S., Engel, M., Sexton, H., et al. (2007). School readiness and later achievement. *Developmental Psychology, 43*(6), 1428–1446.

Fitzmaurice, G. M., Laird, N. M., & Ware, J. H. (2004). *Applied longitudinal analysis.* Hoboken, NJ: Wiley.

Floyd, R. G., Hojnoski, R. L., & Key, J. (2006). Preliminary evidence of technical adequacy of the Preschool Numeracy Indicators. *School Psychology Review, 35,* 627–644.

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

Fuchs, L. S. (2004). The past, present, and future of curriculum-based measurement research. *School Psychology Review, 33,* 188–192.

Fuchs, L. S., &b Deno, S. L. (1991). Paradigmatic distinctions between instructionally relevant measurement models. *Exceptional Children, 57,* 488–500.

Gersten, R., & Chard, D. (1999). Number sense: Rethinking arithmetic instruction for students with mathematical disabilities. *Journal of Special Education, 33,* 18–28.

Ginsburg, A., & Baroody, B. (2003). *Test of Early Mathematics Ability—Third Edition.* Wood Dale: IL: Stoelting.

Greenwood. C. R., Walker, D., Carta, J. J., & Higgins, S. K. (2006). Developing a general outcome measure of growth in the cognitive abilities of children 1 to 4 years old: The early problem-solving indicator. *School Psychology Review, 35,* 535–551.

Hintze, J. M. (2008). Curriculum-based assessment. In T. Gutkin & C. Reynolds (Eds.), *Handbook of school psychology* (4th ed., pp. 397–409). Hoboken, NJ: Wiley.

Jordan, N. C., Kaplan, D., Locuniak, M. N., & Ramineni, C. (2007). Predicting first-grade math achievement from developmental number sense trajectories. *Learning Disabilities Research and Practice, 22,* 36–46.

Jordan, N. C., Kaplan, D., Olah, L. N., & Locuniak, M. N. (2006). Number sense growth in kindergarten: A longitudinal investigation of children at-risk for mathematics difficulties. *Child Development, 77,* 153–175.

Lembke, E. S., & Foegen, A. (2009). Identifying early numeracy indicators for kindergarten and first-grade students. *Learning Disabilities Research & Practice, 24,* 12–20.

Lembke, E. S., Foegen, A., Whittaker, T., & Hampton, D. (2008). Establishing technically adequate measures of progress in early numeracy. *Assessment for Effective Intervention, 33,* 206–214.

Martinez, R. S., Missall, K. N., Graney, S. B., Aricak, O. T., & Clarke B. (2009). Technical adequacy of early numeracy curriculum-based measurement in kindergarten. *Assessment for Effective Intervention, 34,* 116–125.

McConnell, S. R. (2000). Assessment in early intervention and early childhood special education: Building on the past to project into our future. *Topics in Early Childhood Special Education, 20,* 43–48.

McConnell, S. R., Priest, J. S., Davis, S. D., & McEvoy, M. A. (2002). Best practices in measuring growth and development for preschool children. In A. Thomas & J. Grimes (Eds.), *Best practices in school psychology* (4th ed., Vol. 2, pp. 1231–1246). Washington, DC: National Association of School Psychologists.

Methe, S. A., Hintze, J. M., & Floyd, R. G. (2008). Validity and decision accuracy of early numeracy skill indicators. *School Psychology Review, 37,* 359–373.

Missall, K. N., Reschly, A., Betts, J., McConnell, S., Heistad, D., Pickart, M., et al. (2007). Examination of the predictive validity of preschool early literacy skills. *School Psychology Review, 36,* 433–452.

National Center for Education Statistics. (2007). *The nation's report card: Mathematics* 2007 (NCES Publication No. 2007–494). Washington, DC: U.S. Department of Education.

National Center on Response to Intervention. (n.d.). Evidence-based tools and interventions. Retrieved July 1, 2009, from http://www.rti4success.org/index.php?option=com_content&task=view&id=20&Itemid=71

National Council of Teachers of Mathematics. (2000). *Principles and standards for school mathematics.* Reston, VA: Author.

National Mathematics Advisory Panel. (2008). *Foundations for success: The final report of the National Mathematics Advisory Panel.* Washington, DC: U.S. Department of Education.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Thousand Oaks, CA: Sage.

Reid, E. E., Morgan, P., L., DiPerna, J. C., & Lei, P. (2006). Development of measures to assess young children's early academic skills: Preliminary findings from a Head Start–University Partnership. *Insights on Learning Disabilities, 3,* 25–38.

Sanford, A. R., Zelman, J. G., Hardin, B. J., & Peisner-Feinberg, E. S. (2003). *Learning Accomplishment Profile—Third Edition (LAP-3).* Lewisville, NC: Kaplan Press.

Shinn, M. R. (1989). *Curriculum-based measurement: Assessing special children.* New York: Guilford Press.

Trister Dodge, D., Colker L. J., & Heroman, C. (2002). *The Creative Curriculum for Preschool, Fourth Edition.* Washington, DC: Teaching Strategies, Inc.

VanDerHeyden, A. M. (2005). Intervention-driven assessment practices in early childhood/early intervention: Measuring what is possible instead of what is present. *Journal of Early Intervention, 28,* 28–33.

VanDerHeyden, A. M., Broussard, C., Cooley, A. (2006). Further development of measures of early math performance for preschoolers. *Journal of School Psychology*, *44*, 533–553.

VanDerHeyden, A. M., Broussard, C., Fabre, M., Stanley, J., Legendre, J., & Creppell, R. (2004). Development and validation of curriculum-based measures of math performance for preschool children. *Journal of Early Intervention, 27,* 27–41.

VanDerHeyden, A. M., & Snyder, P. (2006). Integrating frameworks from early childhood intervention and school psychology to accelerate growth for all young children. *School Psychology Review, 35,* 519–534.

VanDerHeyden, A. M., Witt, J. C., Naquin, 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., & Mather, N. (2001). *Woodcock-Johnson III Tests of Achievement.* Itasca, IL: Riverside Publishing.

Date Received: January 5, 2009
Date Accepted: July 18, 2009
Action Editor: Sandra Chafouleas ■

Robin Hojnoski is Assistant Professor in the School Psychology Program at Lehigh University. She is interested in the application of school psychology principles and practices to early education. Her research focuses on assessment and intervention to improve educational and social outcomes for preschool children and their families.

Benjamin Silberglitt completed his PhD in Educational Psychology at the University of Minnesota—Twin Cities. His areas of interest include student assessment, school organization and systems-level change, problem solving, response to intervention, and data-based decision making. He is Director of Software Applications for TIES, a technology cooperative of 39 Minnesota school districts.

Randy G. Floyd is Associate Professor of Psychology at The University of Memphis. He received his doctoral degree in school psychology from Indiana State University. His research interests include the structure, measurement, and correlates of cognitive abilities; the technical properties of early numeracy measures; and the process of professional publication.