School Psychology Review,
2006, Volume 35, No. 4, pp. 627–644

GENERAL ARTICLE RELATED TO SPECIAL SERIES

Preliminary Evidence of the Technical Adequacy of the Preschool Numeracy Indicators

Randy G. Floyd, Robin Hojnoski, and Jennifer Key
The University of Memphis

Abstract. Given the potential long-term outcomes of poorly developed mathematical skills, there is a need to understand the emergence and development of these skills in the context of improving educational experiences and ensuring better outcomes for preschoolers. The purpose of this article is to describe preliminary reliability and validity evidence supporting the Preschool Numeracy Indicators (PNIs), which are tasks for the assessment of number skills during the preschool years. Results from a sample of 3- to 6-year-old children (N = 163) attending preschool and Head Start programs revealed (a) evidence of test–retest reliability, (b) evidence that the four PNIs measure the same construct, (c) evidence of age-related differences in PNIs scores, and (d) evidence that the PNIs demonstrate external relations with scores from three test batteries measuring mathematics and number skills and other preacademic skills. Implications focus on the use of the PNIs for screening purposes, and future research is described.

This article includes results based on data from the third author's dissertation research. Portions of this research were presented at the annual meeting of the National Association of School Psychologists (2006).

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 for the contributions of Kelly Harrigan, who constructed task materials and entered the data used in this study. Amberly Barry, Kelly Harrigan, Liz Ingram, Chad Krisak, Allison Margulies, and Jayme McIntyre contributed to the development of the task materials, and Jason Johnson and Tammy McClure contributed to the testing of Sample 4. Tim Keith provided valuable advice about the assumptions and interpretation of fit statistics from the confirmatory factor analysis. We also recognize the influences of the following scholars who discussed issues related to measurement of early numeracy with us through the Early Numeracy Research Group: Ben Clarke, John Hintze, Rebecca Martinez, William Mathews, Scott Methe, Edward Shapiro, Mark Shinn, and Amanda VanDerHeyden.

Correspondence regarding this article should be addressed to Randy G. Floyd, The University of Memphis, Department of Psychology, Memphis, TN 38152; E-mail: rgfloyd@memphis.edu

The general outcome measurement (GOM) approach includes methods for formative evaluation of student progress toward long-term goals through the use of reliable and valid measures of targeted skills tied to the curriculum and representing valued educational goals (Fuchs & Deno, 1991). Perhaps the most well-researched GOM system is curriculum-based measurement (CBM; Deno, 1985). CBM assessment tasks include those targeting general reading achievement, reading comprehension, mathematics computation, spelling, and written expression, and there is a sizable body of reliability and validity evidence supporting the use and interpretation of most of these measures as well as evidence supporting their utility in school settings (see Shinn, 1989, 1998; Shapiro, 2005).

Because of the research supporting the use of CBM with school-age children, additional research has been directed at developing similar measurement systems for younger children. Such systems would allow for monitoring of skill development at the earliest points possible and facilitate timely early intervention to prevent learning difficulties. Most notably, the University of Kansas, the University of Minnesota, and the University of Oregon collaborated to produce a measurement system for infants, toddlers, preschoolers, and early elementary students (Early Childhood Research Institute on Measuring Growth and Development, 1998a, 1998b). Currently, this group has developed measures for several domains, including expressive communication (Luze et al., 2001), social skills (Carta, Greenwood, Luze, Cline, & Kuntz, 2003), early literacy (Good & Kaminski, 2002; McConnell, Priest, Davis, & McEvoy, 2002), and movement (Greenwood, Luze, Cline, Kuntz, & Leitschuh, 2002).

GOMs for Mathematics and Number Skills

There is a need to understand more fully the emergence and growth of mathematics and number skills as young children develop in order to improve educational experiences and to ensure better outcomes for all children (Gersten & Jordan, 2005; Jitendra, 2005). Despite this need, research examining assessment tasks focusing on mathematics and number skills has been limited, even for CBM (Shapiro, 2005). Research including such tasks that target preschoolers has been even more limited. To provide a summary of the existing assessment tasks developed from a GOM framework that tap into mathematics and number skills and to gain insights into measurement of such skills in preschoolers, a comprehensive literature search was conducted to identify peer-reviewed journal articles that have included such tasks (Cooper, 1998; Lipsey & Wilson, 2001). Based on this search, 18 peer-reviewed, published studies were identified that included mathematics and number skills tasks that were either (a) the foci of the study or (b) used as GOMs in intervention research.[1] Of these 18 studies, 11 included only children who were of school age (Grade 1 and higher). Across these 11 studies, there was little variability in the types of tasks used. Notably, 8 of these studies employed tasks requiring children to complete worksheets of mathematics calculation problems of like or varying operations. Furthermore, 5 of these studies employed tasks requiring mathematics applications, mathematics concepts, or both. These tasks frequently included "story problems," or they involved reasoning with numbers, such as finding patterns in lists of numbers.

Although a sizable research base focusing on CBM measures of mathematics with school-age children emerged and review of the descriptions of the task was informative, the types of tasks used with school-age children did not appear directly applicable to preschoolers for at least two reasons. First, most preschoolers do not yet have calculation skills, and their number skills may reflect only an intuitive sense (Clements, 1999). Second, the tasks just described were administered in groups in all but one study, whereas it is generally recommended that preschoolers complete academic tasks on an individual basis (e.g., Bracken, 2000; Magliocca, Rinaldi, Crew, & Kunzelmann, 1977).

The literature search also revealed 7 published studies that include tasks measuring component skills of mathematics, such as counting and number skills, that are potentially more applicable to young children (Chard et al., 2005; Clarke & Shinn, 2004; Daly, Wright, Kelly, & Martens, 1997; Joyce & Wolking, 1987; Magliocca et al., 1977; VanDerHeyden et al., 2004; VanDerHeyden, Witt, Naquin, & Noell, 2001). All studies included preschoolers, children in first grade, or both, but only 1 study included children as young as age 3 (i.e., Magliocca et al., 1977). In contrast to the three types of tasks used with school-age children, these 7 studies included tasks varying widely in their administration formats, stimulus materials, and response requirements. However, three types of tasks appeared to be most common and most well supported by evidence of technical adequacy. Tasks requiring children to count forward orally from the number 1 were included in 4 studies (Chard et al., 2005; Clarke & Shinn, 2004; Daly et al., 1997; VanDerHeyden et al., 2004). These oral counting tasks demonstrated evidence of test–retest reliability across periods from 1 day to approximately 6 months (r = .71–.88; Clarke & Shinn, 2004; Daly et al., 1997; VanDerHeyden et al., 2004). All 4 studies reported evidence of positive relations with concurrently obtained informant-based and test-based measures of mathematics and number skills (r = .12–.91).

Tasks requiring children to identify numerals also appear to have evidence of technical adequacy with preschoolers. Tasks requiring naming of numerals were included in 5 studies, and tasks requiring selection of numerals that had been named orally by an examiner were included in two studies (Chard et al., 2005; Clarke & Shinn, 2004; Daly et al., 1997; Joyce & Wolking, 1987; Magliocca et al., 1977; VanDerHeyden et al., 2004). These numeral identification tasks demonstrated alternate-form reliability (r = .83–.93; Clarke & Shinn, 2004; VanDerHeyden et al., 2004) as well as test–retest reliability across periods from 2 weeks to approximately 6 months (r = .67–.85; Clarke & Shinn, 2004; Daly et al., 1997). Evidence of validity was apparent from their typically moderate relations with other measures of mathematics and number skills (r = .03–.70; Chard et al., 2005; Clarke & Shinn, 2004; VanDerHeyden et al., 2004; cf. Daly et al., 1997).[2] Finally, a third type of task with evidence for its use required children to count objects or images. Such counting tasks were included in 4 studies (Joyce & Wolking, 1987; Magliocca et al., 1977; VanDerHeyden et al., 2004; VanDerHeyden et al., 2001). Most notably, counting tasks developed by VanDerHeyden et al. (2004) and VanDerHeyden et al. (2001) demonstrated evidence of alternate-form reliability (r = .81–.87) and positive relations with other measures of school readiness, mathematics, and number skills (r = .29–.61).

Purpose of the Study

Based on our review of the research, we believe evidence suggests that mathematics and number skills can be reliably and validly represented by tasks for preschoolers developed in the GOM tradition. Consistent with the goals and use of GOMs, in general, and with research supporting the use and interpretation of such measures of early literacy and language with preschoolers (e.g., Good & Kaminski, 2002, and McConnell et al., 2002), we also believe (a) the assessment of mathematics and number skills should begin as early as age 3 to promote the early identification of children at risk and (b) the same assessment tasks should be used with 3- to 6-year-old children to promote optimal comparisons across the preschool years using common measures of these skills.

Building on these beliefs, this study was intended to extend the existing research in several ways. First, the full age range of preschoolers was deliberately included. Second, tasks not evaluated in much previous research but important to mathematical development were adapted or developed. Third, careful attention was paid to the design and use of these tasks to reduce influences that may undermine accurate measurement of mathematics competencies. In summary, in a manner consistent with the first stage in programmatic research

Table 1
Participant Demographic Characteristics (in Percentages) for the Four Subsamples and the Total Sample

	Sample				
Characteristic	1 (n = 47) %	2 (n = 51) %	3 (n = 22) %	4 (n = 43) %	Total (N = 163) %
Age, years					
3	21	16	41	2	17
4	34	35	36	37	36
5	24	49	18	47	37
6	21	—	5	14	10
Sex					
Girl	55	53	50	44	51
Boy	45	47	50	56	49
Race–Ethnicity					
African American	—	100	14	91	57
White	96	—	68	—	34
Hispanic	—	—	5	5	2
Other	4	—	14	5	4
Disability status	0	[a]	9	19	11[a]

[a]Data reflecting disability status were unavailable for Sample 2.

on tasks in the GOM tradition (Fuchs, 2004), this research was intended to generate preliminary evidence supporting the use and interpretation for four indicators of number skills that are collectively named the Preschool Numeracy Indicators (PNIs).

Although issues of sensitivity to growth over time and sensitivity to the effects of instruction are central to the development of GOMs, it is necessary to first examine the technical features of scores from one point in time to ensure they can be interpreted as meaningful (Fuchs, 2004; Fuchs & Fuchs, 1999). Thus, as an initial step in developing GOMs for preschoolers focused on number skills and in a manner consistent with the *Standards for Educational and Psychological Testing* (American Educational Research Association [AERA], American Psychological Association [APA], & National Council on Measurement in Education [NCME], 1999), we examined the technical features of the PNIs in terms of test–retest reliability and four types of validity evidence—content, response processes, internal relations, and external relations with age and with scores from other assessment instruments. Technical adequacy across these features would most directly support the PNIs as screening measures.

Method

Participants and Settings

As shown in Table 1, 163 children ages 3–6 years old participated in this research (M = 56.9 months, SD = 9.7 months).[3] Children were recruited from four preschool education settings to promote the generalization of this study's results to children from a variety of preschool experiences and cultural backgrounds.

Table 1 lists some of the demographic characteristics of the children from each sample. Children in Sample 1 attended a church-affiliated center accredited by National Asso-

ciation for the Education of the Young Child that serves children between the ages of 3 and 6 in classrooms grouped according to age. The center serves families of middle to high socioeconomic status. Children in Sample 2 and Sample 4 attended Head Start programs. The Head Start programs follow Head Start federal regulation curriculum guidelines. The centers serve families of low socioeconomic status. Children in Sample 3 attended a university-affiliated center accredited by National Association for the Education of the Young Child. The center serves families of middle to high socioeconomic status.

Measures

PNIs. To develop tasks yielding GOMs appropriate for representing the mathematics and number skills of preschoolers, a number of steps were taken to provide validity evidence based on content. First, seminal texts focusing on the cognitive development of number skills (e.g., Fuson, 1988; Geary, 1994) and standards for early childhood mathematics education and Head Start (e.g., Clements, Sarama, & DiBiase, 2004; National Council of Teachers of Mathematics, 2000) were reviewed. These reviews led us to target *number and operations* as the most foundational of the five content areas of mathematics for prekindergarten to Grade 2. (These areas include number and operations, geometry, measurement, algebra, and data analysis and probability [National Council of Teachers of Mathematics, 2000].) Second, the targeted item requirements, item stimuli, response requirements, and possible construct-irrelevant influences were reviewed item by item from relevant subtests of published norm-referenced test batteries.[4] Other experimental assessment tasks and GOM probes measuring mathematics and number skills (e.g., Okamoto & Case, 1996) as well as GOM probes measuring early literacy and language (e.g., Good & Kaminski, 2002; McConnell et al., 2002) were reviewed. Design considerations for the PNIs were also given to (a) ensuring hallmark characteristics of GOMs, such as being efficient in administration and scoring and being capable of repeated administration (McConnell et al., 2002); and (b) reducing possible construct-irrelevant influences stemming from the design of the PNIs and their administration to preschoolers (see Bracken, 2000). To ensure that each task measured children's fluency with numbers, number concepts, and counting, several steps were taken. All tasks were developed to be individually administered. Response requirements were selected to reduce motor requirements (e.g., pencil use) and linguistic requirements. Demonstration and sample items were developed to model and practice the basic requirements of the tasks to prevent such problems as floor effects. In addition, stimulus materials were constructed to reduce within-task visual distractions and opportunities for irrelevant within-task child–examiner conversations. From this synthesis of information, the following tasks were developed.[5]

One-to-One Correspondence Counting Fluency targets the ability to count objects fluently. It requires children to point to and count circles approximately 1 inch in diameter printed on an 8.5- × 11-inch page. The examiner models the counting task with four circles, and the child is asked to count the four circles independently. For the actual task, circles are presented in four rows of five circles. Children have a maximum of 30 s to count all 20 circles. If a child makes an error, either by counting the same circle twice or by counting out of sequence, the last circle counted correctly in sequence from 1 is recorded. If a child pauses for 3 s while counting, the examiner repeats the last number the child stated and says, "Keep counting." 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.

One-to-One Correspondence Counting Fluency is similar to a task used by Magliocca et al. (1977) and VanDerHeyden et al. (2004). However, both sets of researchers used a variety of pictures (e.g., frogs, cats) or a variety of geometric shapes rather than a single type of image, and both included 10 or fewer stimuli on each card across multiple cards. We believed

that using circles would minimize distractions potentially elicited by pictures and that the expansion of the number of items to be counted was necessary for the task to be used with older children. Furthermore, to increase ease of administration, the PNI task employs only a single stimulus page rather than multiple cards.

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. Examiners record the last number stated correctly in sequence from 1 on a scoring sheet that includes numerals 1–100. If a child counts out of sequence, the last number stated correctly is marked. If a child pauses for 3 s while counting, the examiner repeats the last number the child stated and says, "Keep counting." The Fluency score represents the number of numbers stated correctly in sequence from 1 in 1 min.[6] Oral Counting Fluency is almost identical to tasks used by Clarke and Shinn (2004), Daly et al. (1997), and VanDerHeyden et al. (2004), but Daly et al. recorded the number of correct number sequences, rather than numbers correct. Chard et al. (2005) investigated a similar oral counting task, but they required children to count to only 20 and appear to have recorded only the number of correct number sequences.

Number Naming Fluency targets the ability to name numerals fluently. It requires children to say the names of the numerals 0–20 as they are presented one at a time on 8.5- × 11-inch pages. Examiners present these pages in rapid succession for 1 min, and children have 3 s to respond to each page. To prevent repetition of numerals across consecutive pages, the 21 numerals included in the task are presented in random sequences in three complete sets for a total of 63 presentations. Examiners record responses to each numeral. If a child incorrectly names the numerals, no corrective feedback is provided. If the child does not respond within 3 s, the next numeral is presented. The fluency score represents the number of numerals named in 1 min. Number Naming Fluency is similar to tasks used by Magliocca et al. (1977), Daly et al. (1997), Clarke and Shinn (2004), and Chard et al. (2005) in that numerals 0–20 (or 1–20) are sampled and that the task is administered for 1 min. (Chard et al., 2005, required children in kindergarten to identify only numerals 1–10.)

As opposed to the tasks used in these previous studies, numerals in Number Naming Fluency are presented singly as opposed to in a grid. VanDerHeyden et al. (2004) used a 1-min task with a singular presentation of numerals as well, but the task sampled only numerals 1–10.

Quantity Comparison Fluency targets the ability to make judgments fluently about differences in the quantity of object groups. It requires children to identify which side of an 8.5- × 11-inch page contains more circles. Children respond by touching the side of the page with more circles. Each side contains 1–6 circles, and each quantity of circles is represented in a standard fashion across pages. Children complete two samples, and corrective feedback is provided if needed. For example, to facilitate measurement of the ability to compare quantities of objects (and not the ability to count circles), children are discouraged from counting circles and prompted to make judgments based on visual inspection of the two sets of circles on each page. For the actual task, 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. Examiners record responses to each page. No corrective feedback is provided. If the child simply states the larger number of circles, the response is recorded as correct. 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). The conceptual demand of Quantity Comparison Fluency is similar to that of tasks used by Clarke and Shinn (2004) and Chard et al. (2005). However, these two studies used numerals from 0–20 and 1–20, respectively, rather than circles, and they presented the numerals in a grid rather than including only one item per page. (Chard et al., 2005, required

children in kindergarten to complete the task using only numerals 1–10.)

Bracken Basic Concepts Scale—Revised (BBCS-R). The BBCS-R (Bracken, 1998) measures young children's understanding of basic concepts. Its first 6 subtests yield the school readiness composite (SRC) and include tasks focused on number, color, letter, and size concepts. The SRC is represented by standard scores with a mean of 100 and a standard deviation of 15. The median internal consistency reliability of the SRC for children 3–5 years old is .96. Test–retest reliability for the SRC is .85 (Bracken, 1998). The BBCS-R Quantitative subtest measures knowledge of foundational concepts related to quantity. It is not included in the SRC. The subtest produces standard scores with a mean of 10 and a standard deviation of 3. The median internal consistency reliability of the Quantitative subtest for children 3–5 years old is .95. Test–retest reliability is .73. Evidence based on content, internal relations, and external relations support the validity of the SRC and Quantitative subtests. For example, the correlation between the SRC and the Wechsler Preschool and Primary Scale of Intelligence—Revised Full Scale IQ (Wechsler, 1989) is .86 (Bracken, 1998).

Woodcock–Johnson III (WJIII). The Applied Problems test from the WJ III Tests of Achievement (Woodcock, McGrew, & Mather, 2001) measures math reasoning skills. The test requires young children to answer mathematical questions that involve using fingers to represent quantity, counting objects, and comparing quantities. The Applied Problems test produces standard scores with a mean of 100 and a standard deviation of 15. The median internal consistency reliability of the Applied Problems test for children 3–5 years old is .92, and 1-year test–retest reliability for children 2–7 years old is .90. Evidence based on internal relations and external relations support its validity. For example, Applied Problems demonstrated sizable factor coefficients on factors measuring quantitative knowledge and fluid reasoning (McGrew & Woodcock, 2001).

Test of Early Mathematics Ability—Third Edition (TEMA-3). The TEMA-3 (Ginsburg & Baroody, 2003) contains four categories of items that assess informal mathematics: (a) numbering skills, (b) number-comparison facility, (c) calculation skills, and (d) understanding of concepts. The TEMA-3 has two parallel forms, Form A and Form B. Both produce a score with a mean of 100 and a standard deviation of 15. Alternate-form reliability for the TEMA-3 is .93. The median internal consistency reliability of the TEMA-3 for children 3–5 years old is .93 for Form A and .95 for Form B. Two-week test–retest reliability is .82 for Form A and .93 for Form B. TEMA-3 has demonstrated moderate to strong correlations with several measures of mathematics and number skills. For example, the correlation between the TEMA-3 score and the WJ III Applied Problems test is .55 (Ginsburg & Baroody, 2003).

Procedures

All children were recruited through letters of invitation sent to parents or guardians through their classroom teachers. Parents returned signed consent forms and provided demographic information about their children. Testing was conducted on an individual basis in classrooms or in spaces close to the classrooms. All PNIs were administered by the three authors in a standard order: One-to-One Correspondence Counting Fluency, Oral Counting Fluency, Number Naming Fluency, and Quantity Comparison Fluency. Testing for Samples 1–3 was conducted in April and May. Testing for Sample 4 was conducted in June.

To provide an estimate of test–retest reliability, 45 children from Sample 1 (94.6%) completed all four PNIs across a 2- to 4-week interval (M = 26.5 days, SD = 3.0 days, range = 13–34 days). From this subsample, 21 children (46.6%) completed the PNIs twice with the same examiner, and 24 children (53.4%) completed the PNIs with different examiners for the initial testing and for the follow-up testing. To provide an estimate of

Table 2
Descriptive Statistics and Correlations Between the PNIs Measures for the Total Sample

Measure	Descriptive Statistics				Correlations			
	M	*SD*	S	K	1	2	3	4
1. One-to-One Correspondence Counting Fluency	29.3	25.1	0.5	1.2	—	.75	.68	.63
2. Oral Counting Fluency	31.0	24.5	1.3	1.0	.58	—	.80	.63
3. Number Naming Fluency	13.3	12.7	1.4	1.0	.47	.65	—	.62
4. Quantity Comparison Fluency	17.9	10.7	−0.2	−0.9	.45	.43	.42	—

Note. PNIs = Preschool Numeracy Indicators; S = skewness; K = kurtosis. Pearson product-moment correlation coefficients between measures are presented above the diagonal (N = 163), and partial correlations between measures with the effects of age removed are reported below the diagonal (N = 162). All correlations are statistically significant, $p < .05$, two tailed.

test–retest reliability over a longer period, all children from Sample 4 completed the four PNIs across a 5- to 7-week interval (M = 48.0 days, SD = 1.7 days, range = 41–52 days). From this subsample, 25 children (58.1%) completed the PNIs twice with the same examiner, and 18 children (41.9%) completed the PNIs with different examiners.

To provide evidence of external relations with like tasks, 21 children from Sample 2 (41%) and 20 children from Sample 3 (95%) completed the BBCS-R (Bracken, 1998) and the Applied Problems test from the WJ III (Woodcock et al., 2001). The BBCS-R and the WJ III Applied Problems test were administered, in counterbalanced order, by two authors 4 days, on average, after the PNIs (M = 3.6 days, SD = 3.2 days, range = 0–14 days). The subsample of 41 children included approximately equal numbers of children from each age level: twelve 3-year-olds (29.3%), thirteen 4-year-olds (31.7%), and sixteen 5-year-olds (39%). Of this subsample, 22 were girls (53.7%), 19 were boys (46.3%), 24 children were African American (58.5%), 13 were White (31.7%), and 1 was Hispanic (2.4%). To provide additional evidence of external relations, 43 children from Sample 4 completed the TEMA-3 (Ginsburg & Baroody, 2003). The TEMA-3 was administered by two authors and a trained graduate student within 2 days after the PNIs. Of this subsample, 24 (55.8%) completed Form A and 19 (44.2%) completed Form B. The subsample of 43 children included five 3-year-olds (11.6%), twenty-six 4-year-olds (60.5%), and twelve 5-year-olds (27.9%). When sex was considered, 19 children were girls (44.2%), 24 were boys (55.8%), 39 children were African American (90.7%), and 2 were Hispanic (4.7%).

Results

Data screening procedures were conducted using both the total sample of children pooled across sites and the subsamples by age level. Assumptions regarding multivariate normality, absence of outliers, linearity, and homogeneity of variance–covariance matrices were not violated. As evident in Table 2, some PNI variables demonstrated somewhat excessive positive skewness, kurtosis, or both (with values >1.0) for the total sample. Variations from normality were also apparent at each age group. Methods to compensate for the variations from normality and assumptions for more specific analyses are noted in what follows (Tabachnick & Fidell, 2006).

Test–Retest Reliability

Table 3 presents the means and standard deviations from the initial testing and from the follow-up testing for the two samples. Pearson correlation coefficients representing test–retest

Table 3
Descriptive Statistics and Test–Retest Reliability Coefficients for the PNIs Measures

Measure	Initial Testing *M*	Initial Testing *SD*	Follow-Up Testing *M*	Follow-Up Testing *SD*	Pearson Correlation Coefficient	Partial Correlation Coefficient (With Age in Months Removed)
Sample 1 (n = 45)[a]						
One-to-One Correspondence Counting Fluency	43.0	26.8	37.6	30.7	.62	.32
Oral Counting Fluency	46.9	4.8	39.6	4.2	.90	.71
Number Naming Fluency	22.2	15.3	24.8	16.1	.91	.73
Quantity Comparison Fluency	24.7	8.6	30.9	10.7	.89	.72
Sample 4 (n = 17)						
One-to-One Correspondence Counting Fluency	23.5	20.1	22.7	21.6	.96	.92
Oral Counting Fluency	23.8	16.5	23.5	21.3	.83	.70
Number Naming Fluency	12.1	13.5	11.9	11.3	.88	.81
Quantity Comparison Fluency	14.5	10.8	14.8	11.3	.94	.86

Note. PNIs = Preschool Numeracy Indicators. All correlation coefficients are statistically significant, $p < .05$, two tailed.
[a]Age in months was unavailable for one child, so n = 44 for the partial correlations.

reliability for Oral Counting Fluency, Number Naming Fluency, and Quantity Comparison Fluency are strong to very strong across both samples, and all exceed the minimal standard (.80) for the reliability of individual measures in psychology and education (e.g., Bracken, 1987). Only One-to-One Correspondence Counting Fluency demonstrated a test–retest reliability coefficient that is below the minimal standard in Sample 1 (r = .62). However, that coefficient was very strong in Sample 4 (r = .96). As evident in Table 3, when the effect of age (in months) was removed via partial correlations, all test–retest reliability coefficients, except for One-to-One Correspondence Counting Fluency in Sample 1, were strong (.70 or greater).[7]

Internal Relations

Table 2 presents the means and standard deviations for the four PNIs for the total sample. The upper half of the right side of Table 2 includes the Pearson product-moment correlation coefficients between the four PNIs. All were moderate to strong.[8] Because some PNIs demonstrated notable deviations from normality, correlations in Table 2 were compared to those from another analysis that included scores from One-to-One Correspondence Counting Fluency, Oral Counting Fluency, and Number Naming Fluency that were transformed (using the Log10 algorithm) to produce more normal distributions. The correlations using these transformed variables were very similar in magnitude and pattern, so the results using the original variables are reported and maintained through the additional analysis of internal structure. In the next step, the relations between the PNIs with the effects of age (in months) removed were calculated using partial correlations. These partial correlations are presented in the lower half of the right side of Table 2. Even with the effects of age removed, all correlations remained in the moderate range.

To provide more information about the relations among the four PNIs, Amos 5.0 (Arbuckle & Wothke, 2004) was used to conduct

Table 4
Fit Statistics for the PNIs Models

Model	χ^2	*df*	*p*	TLI	CFI	Standardized RMR	RMSEA (90% Interval)	AIC
Single-Factor Model	4.726	2	.084	.978	.993	.0189	.096 (.000–.205)	20.956
Two-Factor Model 1 (F1: OOCCF and QCF, F2: OCF and NNF)	1.358	1	.244	.995	.999	.0085	.047 (.000–.221)	19.358
Two-Factor Model 2 (F1: OOCCF and OCF, F2: NNF and QCF)	4.923	1	.027	.942	.990	.0192	.156 (.043–.304)	22.923
Two-Factor Model 3 (F1: OOCCF and NNF, F2: OCF and QCF)	Not admissible							

Note. PNIs = Preschool Numeracy Indicators; F = factor; OOCCF = One-to-One Correspondence Counting Fluency; OCF = Oral Counting Fluency; NNF = Number Naming Fluency; QCF = Quantity Comparison Fluency; CFI = comparative fit index; TLI = Tucker–Lewis index; standardized RMR = standardized root mean square residual; RMSEA = root mean square error of approximation; AIC = Akaike information criterion.

confirmatory factor analysis. Maximum likelihood estimation was used to estimate free parameters in the model to test the fit of the model to the data. Means, standard deviations, and unadjusted Pearson correlations reported in Table 2 were used as input. A *model comparison approach* was used to determine the best-fitting model. The first model included a single factor hypothesized to affect all four PNIs (the Single-Factor Model). This factor was labeled Number Sense (Chard et al., 2005). Results revealed that all paths from the Number Sense factor to the PNIs task scores were statistically significant ($p < .05$). The first row of Table 4 presents relevant fit statistics for this model. The Tucker–Lewis index, the comparative fit index, and the standardized root mean square residual indicate that the model provides good fit to the data (Hu & Bentler, 1999), but the root mean square error of approximation was notably greater than .05, which indicates mediocre fit.

To investigate alternate models specifying more specific factors affecting the PNIs, 3 two-factor models, in which each factor affected two PNIs measures, were developed and fitted (see Table 4). Results revealed that the Two-Factor Model 3 was inadmissible because the relations between the two factors were greater than unity, as revealed through a nonpositive definite covariance matrix. Both the Two-Factor Model 1 and the Two-Factor Model 2 revealed that their two factors were highly intercorrelated, $r = .94$ and $r = .99$, respectively. Fit statistics in Table 4, with particular emphasis on the Akaike information criterion values, indicated that the Two-Factor Model 1 provided a somewhat better fit than Two-Factor Model 2. (Lower Akaike information criterion values indicate better-fitting hypothesized models.) Comparisons of the χ^2 values and the degrees of freedom from the Single Factor Model and the Two-Factor Model 1 revealed that the Two-Factor Model 1 did not provide significantly better fit than the Single Factor Model ($\Delta\chi^2 = 3.37$, $\Delta df = 1$, $p > .05$).[9] Therefore, the latent structure most parsimoniously describing the relations between the PNIs includes only a single factor.

External Relations

Several analyses were completed to examine the relations between select external

Table 5
Descriptive Statistics and Results of the Cross-Sectional Analysis by PNIs Measure

	Age Level							
	3 (n = 28)		4 (n = 58)		5 (n = 60)			
Measure	M	SD	M	SD	M	SD	ANOVA $F(2, 143)$	Effect Size η^2
One-to-One Correspondence Counting Fluency	15.8	17.7	19.7	20.2	38.0	24.0	15.1*	.25
Oral Counting Fluency	16.4	12.9	22.2	16.9	37.7	23.4	15.9*	.30
Number Naming Fluency	6.0	4.5	8.5	7.8	16.4	12.5	15.2*	.34
Quantity Comparison Fluency	12.7	9.2	14.7	10.0	21.2	9.6	10.0*	.17

Note. PNIs = Preschool Numeracy Indicators; ANOVA = analysis of variance.
*$p < .05$.

variables and the PNIs scores and to examine the effects of such variables on these scores. These external variables were age and measures of early academic skills.

Age. To examine the relations between the PNIs and the external variable age, two analyses were completed using the total sample. First, Pearson product-moment correlation coefficients were calculated between age in months and scores from each task. Age correlated positively and moderately (r = .62 for One-to-One Correspondence Counting Fluency, r = .68 for Oral Counting Fluency, r = .66 for Number Naming Fluency, and r = .54 for Quantity Comparison Fluency). These correlations using the original variables were not notably different from those using the transformed variables for One-to-One Correspondence Counting Fluency, Oral Counting Fluency, and Number Naming Fluency.

Although evidence based on internal relations supports the assertion that the PNIs measure the same latent construct, the focus of this article is on the technical features of the scores from each task. Therefore, in a second analysis examining relations with age, four one-way analyses of variance were completed for each PNI in which age (in years) was used as the independent variable. Because the subsample of 6-year-old children was small (n = 17), each analysis focused on three age levels: 3-year-olds (n = 28), 4-year-olds (n = 58), and 5-year-olds (n = 60). In addition to meeting the assumptions described earlier, there was homogeneity of variance across age groups ($F_{max} < 10$). As evident in Table 5, age demonstrated significant main effects for all four measures, $p < .05$. Post hoc tests (Tukey's honestly significant difference) revealed that, across all PNIs, 5-year-old children obtained significantly higher scores than 3-year-old children and 4-year-old children ($p < .05$ for all pairwise comparisons). However, there was no significant difference between 4-year-old and 3-year-old children. Findings across these analyses remained the same (a) using the transformed variables for One-to-One Correspondence Counting Fluency, Oral Counting Fluency, and Number Naming Fluency and (b) using only the homogeneous Sample 1, for which there were relatively equal sample sizes at each age (see Table 1).

Measures of early academic skills. Pearson product-moment correlation coefficients were calculated between the PNIs measures and the scores from the BBCS-R (Bracken, 1998) and the WJ III Applied Problems test (Woodcock et al., 2001).[10] In Sample 3, there was no significant difference be-

Table 6
Relations Between the PNIs Measures and Scores from the Bracken Basic Concepts Scale—Revised, the Woodcock–Johnson III Applied Problems Test, and the Test of Early Mathematics Ability—Third Edition

			Correlations							
			OOCCF		OCF		NNF		QCF	
Measure	*M*	*SD*	U	C	U	C	U	C	U	C
BBCS-R School Readiness Composite	102.5	16.4	.41*	.38*	.57*	.54*	.43*	.40*	.61*	.58*
BBCS-R Quantitative	9.2	3.6	.36*	.31*	.36*	.31*	.34*	.29*	.52*	.46*
WJ III Applied Problems	96.5	21.6	.40*	.29	.45*	.33*	.49*	.36*	.51*	.38*
TEMA-3	84.0	11.6	.54*	.64*	.55*	.55*	.60*	.70*	.48*	.58*

Note. $N = 41$. BBCS-R = Bracken Basic Concepts Scale—Revised; WJ III = Woodcock–Johnson III; TEMA-3 = Test of Early Mathematics Ability—Third Edition. OOCCF = One-to-One Correspondence Counting Fluency, OCF = Oral Counting Fluency, NNF = Number Naming Fluency, and QCF = Quantity Comparison Fluency. U = uncorrected Pearson product-moment correlation coefficients, C = corrected Pearson product-moment correlation coefficients.
*$p < .05$, two tailed.

tween the TEMA-3 score from Form A and Form B (Ginsburg & Baroody, 2003) and no significant differences in the magnitude of the correlations between Form A scores and each PNI and Form B scores and each PNI. Therefore, Pearson product-moment correlation coefficients were calculated between the TEMA-3 scores, regardless of form, and each PNI measure. For all these analyses, the PNIs were standardized for each whole-year age level by converting raw scores to z scores ($M = 0$, $SD = 1$) for the subsample of 3-, 4-, and 5-year-olds. This data transformation was completed for two reasons. First, the PNIs demonstrate moderate relations with and significant effects of age, as described earlier. Second, the age-based scores (not the raw scores) from the BBCS-R, the WJ III Applied Problems test, and the TEMA-3 have accepted evidence of reliability and validity.

As evident in Table 6, the mean for the BBCS-R SRC is slightly over 0.1 *SD* above the normative average and both the mean for the BBCS-R Quantitative subtest and the mean for the WJ III Applied Problems test are only somewhat below the normative average. These norm-based statistics indicate that the sample included in this analysis is approximately as knowledgeable as the population of preschool-age children. However, there was somewhat more variability in scores than expected. All three standard deviations exceed to some extent the expected standard deviations. In contrast, the mean for the TEMA-3 score from Sample 4 is more than 1 *SD* below the mean, and the variability is less than would be expected.

Because the inflated or restricted variability of these samples' distributions of these external measures affects their correlations with the PNIs, Table 6 presents uncorrected Pearson product-moment correlation coefficients and correlation coefficients corrected for expansion or restriction of range. All corrected coefficients but one were statistically significant ($p < .05$). Only the corrected correlation between One-to-One Correspondence Counting Fluency and the WJ III Applied Problems test was nonsignificant. Of the 16 corrected correlation coefficients, half were moderate in magnitude and the other half were weak in magnitude. Quantity Comparison Fluency demonstrated the highest correlations with the WJ III Applied Problems test and the two BBCS-R measures, whereas Number Naming Fluency demonstrated the highest

correlation with the TEMA-3. When the corrected correlations were converted to z scores and averaged, Quantity Comparison Fluency demonstrated the highest average correlation with the other measures.

Discussion

Based on our belief that the assessment of mathematics and number skills should begin early in development and that these skills should be assessed across the full age range of preschoolers using the same tasks, we developed the PNIs and evaluated their technical features. Drawing from the *Standards for Educational and Psychological Testing* (AERA, APA, & NCME, 1999), this article described aspects of test development and presented reliability and validity evidence for the PNIs using four subsamples that included 163 children from diverse backgrounds.

Development and Evidence Based on Content and Response Processes

The development of the PNIs was supported by validity evidence based on content and by evidence based on response processes (AERA, APA, & NCME, 1999). For example, consistent with the design goal of ensuring that the PNIs could be completed by even the youngest and lowest-functioning children, we found that, out of 163 children who completed the initial assessment with the PNIs, only 4 children (2.5%) obtained scores of 0 on One-to-One Correspondence Counting Fluency, only 2 children (1.2%) obtained a score of 0 on Oral Counting Fluency, only 13 children (8.0%) obtained scores of 0 on Number Naming Fluency, and only 21 children (12.9%) obtained scores of 0 on Quantity Comparison Fluency. Although there was no absence of scores of 0 on any task, the design goal appears to have been met. In addition, because PNI task materials were kept simple (using only numerals or black circles on the stimulus materials), at an informal level of evaluation by the PNI examiners, few task-irrelevant verbalizations from children about these materials were observed.

Reliability

Three of the four PNIs demonstrated test–retest reliability estimates across both 2 to 4 weeks and 5 to 7 weeks that exceed the minimal standard for individual measures in psychology and education (e.g., Bracken, 1987). The consistent test–retest reliability for three of the PNIs indicates that the rank ordering of children's scores from the initial testing to the follow-up testing on each task was relatively similar across brief periods. This consistency is a desirable measurement property because great variation in PNI scores across relatively brief periods would indicate that the targeted skills may not be measured accurately. Test–retest reliability coefficients of similar magnitude to those found in this study have been demonstrated not only with measures of mathematics and number skills with young children (e.g., Clarke & Shinn, 2004; VanDerHeyden et al., 2004) but also with a variety of GOMs with older children (see Marston, 1989). In this study, it is important to note that (a) children who completed the PNIs were not tested under conditions that included instruction addressing number skills, which would likely result in increased scores; and (b) only two performance samples were collected across relatively brief periods, which may be too brief to reflect growth. Therefore, it is likely that the difference in means from initial testing to follow-up testing reflects several sources of measurement error. We believe these sources of error were greater at follow-up testing because of influences such as regression toward the mean, practice effects, and the absence of task novelty affecting speed of performance.

Internal Relations

Consistent with findings from Clarke and Shinn (2004) using scores from three similar tasks, the PNIs demonstrated moderate to strong positive relations between them. Even with the effects of age removed, all correlations remained moderate in magnitude. When confirmatory factor analysis was used to better understand these relations, results revealed that performance on the PNIs is best explained

by a single factor, called Number Sense. Oral Counting Fluency demonstrated the highest relations with this factor, so Oral Counting Fluency can be said to best represent what is common across all four PNIs.

Evidence Based on External Relations

One set of results revealed that all PNIs were positively and moderately correlated with age. Another set of results revealed a sizable effect of age. Post hoc tests indicated that 5-year-old children obtained significantly higher scores than 3-year-old children and 4-year-old children, but significant differences were not apparent when comparing the two youngest age groups. The absence of differences between the two young age groups suggests that there may be sensitive periods in the development of certain mathematical capacities (e.g., during the fifth year). If replicated, the findings could have implications for maximizing instructional efforts through systematic and strategic timing. These findings are also encouraging because they indicate the potential of the PNIs in monitoring development over time. In addition, PNIs scores demonstrated sizable relations with existing measures used to represent school readiness and the early number and mathematical competencies of preschoolers. These results indicate similar patterns of relations as those found between other tasks measuring mathematics and number skills and earlier versions of the criterion measures used in this study (Clarke & Shinn, 2004; Daly et al., 1997; VanDerHeyden et al., 2004).

Which of the PNIs Seems Best?

A challenge in developing GOMs is identifying a measurement task or set of tasks that integrates skills required for successful long-term performance (Fuchs, 2004). This challenge is particularly salient when targeting preschool-age children when skill acquisition can be rapid and quite variable across children. In the interest of parsimony and consistent with GOM systems, the challenge becomes selecting tasks that have demonstrated reliability and validity evidence supporting use of their static scores (Fuchs, 2004).

Some may view Oral Counting Fluency as a rote memorization task, much like reciting the ABCs, and assert that it may provide little evidence of conceptual understanding of numbers, quantities, and relations on the number line (Griffin & Case, 1997). However, the results of this study, as well as those from previous studies using such counting tasks (Aunola, Leskinen, Lerkkanen, & Nurmi, 2006; Chard et al., 2005; Clarke & Shinn, 2004; Daly et al., 1997; VanDerHeyden et al., 2004), indicate that they have the most promise as GOMs with preschoolers and children of early elementary school age. In addition to evidence based on content indicating that oral counting is part of a constellation of important early numeracy skills, in this study, Oral Counting Fluency demonstrated the fewest scores of 0 of all of the PNIs, strong test–retest reliability, and sizable relations with other measures of school readiness and early number and mathematical competencies. Oral Counting Fluency also demonstrated the strongest relations with the Number Sense factor that appears to underlie all of the PNIs. Equal in importance as a potential GOM are that (a) Oral Counting Fluency requires no stimulus materials (in contrast to all other PNIs) and (b) the duration of its administration frequently lasts far less than 1 min per child because the task requires no practice trials and because the task ends when the first counting error is made. Thus, Oral Counting Fluency demonstrates a sizable body of reliability and validity evidence, and it is logically the most efficient and economical to administer. Because all four PNIs were administered in a standard order to all children, additional research will be needed to examine the reliability and validity of scores from Oral Counting Fluency when it is administered in isolation.

Future Directions, Limitations, and Caveats

There is little doubt that additional research should be conducted that (a) targets the use and interpretation of the PNIs using larger,

more nationally representative samples, (b) targets the technical adequacy of the PNIs for each preschool-age level (e.g., 3-year-olds), and (c) is conducted by researchers independent of the PNIs authors. Despite the promising results we have presented, additional research must be conducted to provide evidence that these measures are worthwhile to school psychologists and other professionals.

Administration and scoring accuracy. Information is needed about the accuracy of scoring and administration of the PNIs. Rather than focusing on interscorer (or interobserver) agreement, evidence of scoring accuracy should be obtained by comparing examiner's concurrent scoring of the PNIs to incontrovertible data representing the types of events that should be scored (Johnston & Pennypacker, 1993). These incontrovertible data can be obtained from careful recording of children's responses to the PNIs via video and audio technology (see Daly et al., 1997). In addition, careful recording and coding of examiner behaviors can be use to determine administration fidelity.

External relations. More information about the external relations of the PNIs would be useful. For one, the predictive validity of the PNIs with criterion measures, such as preschool exit scores on norm-referenced published tests and similar GOMs in early elementary school (e.g., Clarke & Shinn, 2004), should be investigated. In addition, it is unknown whether such measures of mathematics and number skills tap into constructs different from language skills and early literacy skills at this age. It is possible that there exists a single factor of academic readiness skills in preschoolers (cf. Shinn, Good, Knutson, Tilly, & Collins, 1992).

Sensitivity to growth and treatment. Although the technical features of performance at one point in time are important in determining the adequacy of measurement tools, for the PNIs to be considered in the general family of GOMs, evidence of sensitivity to growth over time and sensitivity to the effects of instruction or intervention is needed.[11] At present, the PNIs cannot be recommended for progress monitoring. In fact, we know of no such evidence available for supporting the use of GOMs targeting mathematics and number skills with children younger than kindergarten age (cf. Chard et al., 2005; Clarke & Shinn, 2004).

Utility and evidence based on consequences. The utility of the PNIs needs to be evaluated across diverse groups of children. For example, because the preliminary evidence presented here indicates that the PNIs may yield useful screening measures, participants in future research should include children not enrolled in preschool programs. Investigating the treatment utility (aka instructional utility) of the PNIs should also constitute a stage in their evaluation (Fuchs, 2004). Finally, the acceptability of the tasks and their resulting data as well as the extent to which the tasks reflect important outcomes (as judged by parents, teachers, and early childhood administrators) should be examined (Wolf, 1978).

Implications for Practice

Understanding the emergence and growth of mathematics and number skills is important to improve educational experiences and to ensure better outcomes. There is substantial evidence that, as early as 3 years of age, groups of children from high- and low-socioeconomic backgrounds differ substantially in their knowledge and skill development. These differences have been frequently shown to increase through the remainder of the preschool years (Case, Griffin, & Kelly, 1999; Wright, Marland, & Stafford, 2000). There is also evidence that preschool children's mathematics and number skills predict achievement in arithmetic in elementary school and across longer periods in mathematics through Grade 10 (Jimerson, Egeland, & Teo, 1999; Stevenson & Newman, 1986). Early intervention, in general, has been demonstrated to be an effective means of addressing academic skill deficits and improving outcomes related to literacy (Lennon & Slesinski, 1999). With more research to support their

use, the PNIs may offer school psychologists a way to assess the early numeracy development of preschoolers, much like existing tasks commonly used to assess the early literacy development of preschoolers and early elementary school students. The PNIs may provide a means of early identification that decreases the need for later remediation and increases the numbers of children who enter formal schooling with the necessary skills for mathematical success.

Footnotes

[1]A table that includes all of these studies can be obtained from the first author or from visiting http://www.psyc.memphis.edu/images/floyd.shtml

[2]We used the following general labels for r values: *negligible to very weak* (0–.2), *weak* (.2–.4), *moderate* (.4–.7), *strong* (.7–.9), and *very strong* (.9–1.0).

[3] Age in months could not be computed for one 3-year-old child.

[4]This list of assessment tasks can be obtained from the first author.

[5]The PNIs materials can be obtained from contacting the first two authors.

[6]The 60-s metric was included to allow for the possibility that some children might count to 100 in less than 1 min. One child from our sample exceeded 100 in less than 1 min and received a score of 109.

[7]To remove the effects of age on the test–retest correlations, using the larger Sample 1, we also calculated for each measure (a) z scores ($M = 0, SD = 1$) for each 1-year age group at initial testing across the total sample, and (b) z scores for each 1-year age group from Sample 1 at follow-up testing, aggregated the age-adjusted scores across all four age groups, and correlated them. The resulting Pearson correlation coefficients were similar or somewhat higher than those reported via the partial correlations. Spearman rho coefficients for each 1-year age group were also calculated for each measure using Sample 1. These results can be obtained from the first author.

[8]When Pearson product-moment correlation coefficients between PNIs were obtained for each 1-year age group, all coefficients except one (between Oral Counting Fluency and Quantify Comparison Fluency at age 4, $r = .38$, $p < .05$) were positive and moderate in magnitude. These results can be obtained from the first author.

[9]Results supporting the single-factor model were also apparent when age effects were considered. When the means, standard deviations, and partial correlations presented in Table 3, and the means, standard deviations, and correlations resulting from aggregation of z scores for each 1-year age group (at initial testing for the total sample) across all four 1-year age groups were used as input to Amos, no alternate model (a two-factor model) fit significantly better than the single-factor model ($p > .05$). These results can be obtained from the first author.

[10]Because the WJ III Compuscore and Profiles Program (Schrank & Woodcock, 2001) did not yield standard scores for two 3-year-old children who obtained raw scores of 0, these children were assigned standard scores of 55, which is three standard deviations below the population mean.

[11]The first two authors have begun this line of inquiry by collecting monthly data using the PNIs with a sample of children from a Head Start program.

References

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing.* Washington, DC: American Educational Research Association.

Arbuckle, J. L., & Wothke, W. (2004). *Amos 5.0 user's guide.* Chicago: Smallwaters.

Aunola, K., Leskinen, E., Lerkkanen, M. J., & Nurmi, J. E. (2006). Developmental dynamics of math performance from preschool to grade 2. *Journal of Educational Psychology, 96,* 699–713.

Bracken, B. A. (1987). Limitations of preschool instruments and standards for minimal levels of technical adequacy. *Journal of Psychoeducational Assessment, 5,* 313–326.

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

Bracken, B. A. (2000). Maximizing construct relevant assessment: The optimal preschool testing situation. In B. Bracken (Ed.), *The psychoeducational assessment of preschool children* (3rd ed., pp. 33–45). Boston: Allyn & Bacon.

Carta, J. J., Greenwood, C. R., Luze, G. J., Cline, G., & Kuntz, S. (2003). Developing a general outcome measure of growth in social skills for infants and toddlers. *Journal of Early Intervention, 26,* 91–114.

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. J., Clarke, B., Baker, S., Otterstedt, J., Braun, D., & Katz, R. (2005). Using measures of number sense to screen for difficulties in mathematics: Prelim-

inary findings. *Assessment for Effective Intervention, 30,* 3–14.

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

Clements, D. H. (1999). Geometric and spatial thinking in young children. In J. V. Copley (Ed.), *Mathematics in the early years* (pp. 66–79). Washington, DC: National Association for the Education of Young Children.

Clements, D. H., Sarama, J., & DiBiase, A. M. (Eds.). (2004). *Engaging young children in mathematics: Standards for early childhood mathematics education.* Hillsdale, NJ: Erlbaum.

Cooper, H. M. (1998). *Synthesizing research: A guide for literature reviews* (3rd ed.). Thousand Oaks, CA: Sage.

Daly, E. J., III, Wright, J. A., Kelly, S. Q., & Martens, B. K. (1997). Measures of early academic skills: Reliability and validity with a first grade sample. *School Psychology Quarterly, 12,* 268–280.

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

Early Childhood Research Institute on Measuring Growth and Development. (1998a). *Research and development of individual growth and development indicators for children between birth and age eight* (Tech. Rep. No. 4). Minneapolis: University of Minnesota.

Early Childhood Research Institute on Measuring Growth and Development. (1998b). *Theoretical foundations of the Early Childhood Institute on Measuring Growth and Development: An early childhood problem-solving model* (Tech. Rep. No. 6). Minneapolis: University of Minnesota.

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

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

Fuchs, L. S., & Fuchs, D. (1999). Monitoring student progress toward the development of reading competence: A review of three forms of classroom-based assessment. *School Psychology Review, 28,* 659–671.

Fuson, K. C. (1988). *Children's counting and concepts of number.* New York: Springer-Verlag.

Geary, D. C. (1994). *Children's mathematical development: Research and practical applications.* Washington, DC: American Psychological Association.

Gersten, R., & Jordan, N. C. (Eds.). (2005). Early screening and intervention in mathematics difficulties: The need for action [Special issue]. *Journal of Learning Disabilities, 38*(4).

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

Good, R. H., III, & Kaminski, R. A. (Eds.). (2002). *Dynamic Indicators of Basic Early Literacy Skills* (6th ed.). Eugene, OR: Institute for the Development of Educational Achievement.

Greenwood, C. R., Luze, G. J., Cline, G., Kuntz, S., & Leitschuh, C. (2002). Developing a general outcome measure of growth in movement for infants and toddlers. *Topics in Early Childhood Special Education, 22,* 143–157.

Griffin, S., & Case, R. (1997). Re-thinking the primary school math curriculum: An approach based on cognitive science. *Issues in Education, 3,* 1–49.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6,* 1–55.

Jimerson, S., Egeland, B., & Teo, A. (1999). A longitudinal study of achievement trajectories: Factors associated with change. *Journal of Educational Psychology, 91,* 116–126.

Jitendra, A. K. (Ed.). (2005). Mathematics assessment [Special issue]. *Assessment for Effective Intervention, 30*(2).

Johnston, J. M., & Pennypacker, H. S. (1993). *Strategies and tactics of behavioral research* (2nd ed.). Hillsdale, NJ: Erlbaum.

Joyce, B. G., & Wolking, W. D. (1987). Standardized tests and timed curriculum-based assessments: A comparison of two methods for screening high-risk students. *Journal of Psychoeducational Assessment, 5,* 185–193.

Lennon, J. E., & Slesinski, C. (1999). Early intervention in reading: Results of a screening and intervention program for kindergarten students. *School Psychology Review, 25,* 215–227.

Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis.* Thousand Oaks, CA: Sage.

Luze, G. J., Linebarger, D. L., Greenwood, C. R., Carta, J. J., Walker, D., Leitschuh, C., et al. (2001). Developing a general outcome measure of growth in the expressive communication of infants and toddlers. *School Psychology Review, 30,* 383–406.

Magliocca, L. A., Rinaldi, R. T., Crew, J. L., & Kunzelmann, H. P. (1977). Early identification of handicapped children through a frequency sampling technique. *Exceptional Children, 43,* 414–420.

Marston, D. B. (1989). A curriculum-based measurement approach to assessing academic performance: What it is and why do it. In M. R. Shinn (Ed.), *Curriculum-based assessment: Assessing special children* (pp. 18–78). New York: Guilford Press.

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 IV* (pp. 1231–1246). Bethesda, MD: National Association of School Psychologists.

McGrew, K. S., & Woodcock, R. W. (2001). *Technical manual. Woodcock–Johnson III.* Itasca, IL: Riverside Publishing.

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

Okamoto, Y., & Case, R. (1996). Exploring the microstructure of children's central conceptual structures in the domain of number. *Monographs of the Society for Research in Child Development, 61,* 27–59.

Schrank, F. A., & Woodcock, R. W. (2001). WJ III Compuscore and Profiles Program [Computer Software]. *Woodcock–Johnson III.* Itasca, IL: Riverside Publishing.

Shapiro, E. S. (2005). *Academic skills problems: Direct assessment and intervention* (3rd ed.). New York: Guilford Press.

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

Shinn, M. R. (1998). *Advanced applications of curriculum-based measurement.* New York: Guilford Press.

Shinn, M. R., Good, R. H., III, Knutson, N., Tilly, W. D., III, & Collins, V. L. (1992). Curriculum-based measurement of oral reading fluency: A confirmatory analysis of its relation to reading. *School Psychology Review, 21,* 459–479.

Stevenson, H. W., & Newman, R. S. (1986). Long-term predictions of achievement and attitudes in mathematics and reading. *Child Development, 57,* 646–659.

Tabachnick, B. G., & Fidell, L. S. (2006). *Using multivariate statistics* (5th ed.). Needham Heights, MA: Allyn & Bacon.

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., 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.

Wechsler, D. (1989). *Wechsler Preschool and Primary Scale of Intelligence—Revised.* San Antonio, TX: Psychological Corporation.

Wolf, M. M. (1978). The case of subjective measurement or how applied behavior analysis is finding its heart. *Journal of Applied Behavior Analysis, 11,* 203–214.

Woodcock, R. W., McGrew, K. S., & Mather, N. (2001). *Woodcock–Johnson III Tests of Achievement.* Itasca, IL: Riverside Publishing.

Wright, R. J., Marland, J., & Stafford, A. K. (2000). *Early numeracy: Assessment for teaching and intervention.* Thousand Oaks, CA: Paul Chapman.

Date Received: December 23, 2005
Date Accepted: October 3, 2006
Action Editor: James DiPerna ■

Randy G. Floyd is an assistant professor of psychology at the University of Memphis. He received his doctoral degree in school psychology from Indiana State University. His research interests include assessment of cognitive abilities; identification of aptitudes for reading, mathematics, and writing achievement; and improving behavioral assessment methods.

Robin L. Hojnoski is currently an assistant professor of psychology at Lehigh University. She received her doctoral degree in school psychology from the University of Massachusetts Amherst. Her research interests include assessment and intervention for early learning and social–emotional behavior of young children and application of school-based consultation procedures to community early education settings.

Jennifer Key is a doctoral student in the school psychology program at The University of Memphis. She is currently a school psychologist in the Lewisville Independent School District in Lewisville, Texas. Her research interests are in early interventions in mathematics and literacy.