AERA Open
July-September 2016, Vol. 2, No. 3, pp. 1–15
DOI: 10.1177/2332858416653756

Can We Measure the Transition to Reading? General Outcome Measures and Early Literacy Development From Preschool to Early Elementary Grades

Scott McConnell
Alisha Wackerle-Hollman
University of Minnesota

This study evaluated the extent to which existing measures met standards for a continuous suite of general outcome measures (GOMs) assessing children's early literacy from preschool through early elementary school. The study assessed 316 children from age 3 years (2 years prekindergarten) through Grade 2, with 8 to 10 measures of language, alphabetic principle, phonological awareness, and beginning reading. We evaluated measures at each grade group against six standards for GOMs extracted from earlier work. We found that one measure of oral language met five or six standards at all grade levels, and several measures of phonological awareness and alphabetic principle showed promise across all five grade levels. Results are discussed in relation to ongoing research and development of a flexible and seamless system to assess children's academic progress across time for effective prevention and remediation, as well as theoretical and empirical analyses in early literacy, early reading, and GOMs.

Keywords: *early literacy, reading, general outcome measurement, assessment*

Reading proficiency ensures early and continued academic achievement. Research shows that early reading achievement prepares for later learning and achievement (e.g., Chard & Kame'enui, 2000; National Early Literacy Panel [NELP], 2009; National Reading Panel, 2000) and that students struggling to read in the earliest grades are apt to continue struggling through school (cf. Stanovich, 1988). While some historical accounts of factors affecting reading achievement begin analysis with child and instructional practices at or after kindergarten (e.g., Anderson, Hiebert, Scott, & Wilkinson, 1985), theoretical advances (Sénéchal, LeFevre, Smith-Chant, & Colton, 2001; Snow, Burns, & Griffin, 1998; Whitehurst & Lonigan, 1998), descriptive and intervention research (Walker, Greenwood, Hart, & Carta, 1994; Whitehurst et al., 1999), and systematic reviews (Juel, 2006; NELP, 2009; Snow et al., 1998) suggest that children's status at and before kindergarten (K) controls important variance in reading success. This better understanding of children's achievement from preschool to early elementary school can help design and evaluate prevention and early intervention services to avert reading disabilities.

This "age 3 to Grade 3" focus has sparked significant expansion in curricular, programmatic, and other interventions designed to promote early literacy and prevent later reading delays among high-risk children. Federal priorities (e.g., Early Reading First) expand the formal focus of existing programs (i.e., Head Start and initiatives under the Individuals with Disabilities Education Improvement Act of 2004) through Good Start, Grow Smart, and the Preschool for All federal initiatives. States have also acted, increasing the scope of publicly funded early childhood education to include literacy. Commercial publishers have developed and released >20 scientifically based early literacy curricula, and evidence-based practice portals (e.g., What Works Clearinghouse) now review these curricula online.

Theoretical and conceptual models of early literacy development[1] in later preschool (NELP, 2009; Sénéchal et al., 2001; Whitehurst & Lonigan, 1998) relate to, but differ from, theoretical analyses of beginning reading in early elementary school (e.g., Adams, 1990; National Reading Panel, 2000). In general, models of preschool early literacy development emphasize oral language and linguistic processes, phonological sensitivity and analysis, and informal skills for letter-sound correspondence, while models of primary-grade beginning reading emphasize decoding, letter-sound correspondence and other formal phonological analysis tasks, production of oral equivalents of printed words, and proficiency in the production and comprehension of connected text.[2] Little attention goes to the overlaps and boundaries among these models and to ways in which child proficiency in preschool literacy relates to elementary school reading proficiency (McConnell & Wackerle-Hollman, 2006).

Clarifying the overlap between early literacy and beginning reading may be especially important in advancing research and practice that promote proficiency and prevent delays or disabilities throughout the primary grades. Preschool early literacy interventions can promote later reading achievement to the degree that they enrich basic skills for beginning reading (Dickinson & Neuman, 2006; Greenwood, Bradfield, et al., 2011; Snow et al., 1998); a lack of alignment between the outcomes of preschool intervention and required skills for later reading may blunt these relations and may account for noted declines in intervention effects as preschoolers move into elementary school (Cooper & Lanza, 2014). This suggests the importance of a detailed analysis of the relation between early literacy and beginning reading and the identification of measures that describe these relations in ways that will assist in aligning curricula and instruction.

Such an alignment will require improved resources for assessing child performance across the full age range of intervention. Best practices and a growing research base (e.g., Wallace, Espin, McMaster, Deno, & Foegen, 2007) suggest that a coordinated, ongoing assessment system that is both related to long-term outcomes and sensitive to short-term changes in child performance and intervention effects will contribute significantly to improved intervention outcomes. General outcome measures (GOMs; Deno, 1985, 1986, 1997; Deno, Mirkin, & Chiang, 1982; Fuchs & Deno, 1991) in particular can meet the needs of progress monitoring and intervention enhancement in this expanded model. GOMs are brief, easy-to-collect, sensitive, and valid measures of child performance related to long-term outcomes. Research has led to robust procedures and practices for GOMs in elementary and secondary schools, including measures of reading (Wayman, Wallace, Wiley, Tichá, & Espin, 2007), writing (McMaster & Espin, 2007), and mathematics (Foegen, Jiban, & Deno, 2007), and applications to allocate services that promote achievement among children with or at risk for disabilities (Fuchs, 2003; Fuchs & Fuchs, 2007).

The logic of GOMs has helped to develop measures of earlier skills. Dynamic Indicators of Basic Early Literacy Skills (DIBELS) partly represent extending GOMs to reading achievement in the first school years (Good, Gruba, & Kaminski, 2002; Good & Kaminski, 1996; Kaminski & Good, 1996). Individual Growth and Development Indicators (IGDIs) follow similar logic for infants and toddlers (Carta, Greenwood, Luze, Cline, & Kuntz, 2004; Greenwood, Luze, & Carta, 2002; Luze, Linebarger, Greenwood, Carta, & Walker, 2001) and preschoolers (Carta et al., 2005; McConnell & Missall, 2008; McConnell, Priest, Davis, & McEvoy, 2002; Missall, McConnell, & Cadigan, 2006; Priest et al., 2002), with an emphasis on preschool early literacy (Cadigan & Missall, 2007; Hojnoski & Missall, 2006; Missall et al., 2006). To date, these measures have demonstrated correlations with later reading achievement (e.g., Missall et al., 2006) and predictive validity for basic research and intervention evaluation (Missall et al., 2007). Preschool IGDIs and early elementary DIBELS might show promise as part of an aligned portfolio of early literacy and reading assessments (cf. Wallace et al., 2007), but no one has yet evaluated the characteristics of such a portfolio to determine what assessment from preschool early literacy to early elementary beginning reading could be.

To date, most GOMs have been developed and used in ways consistent with classical test theory (Cronbach, 1990); broad and loosely controlled samples of child behavior are summed or scored to serve as measured performance. Given this and based on the recommendations of Deno (e.g., 1997; Deno et al., 1982) and others (e.g., McMaster & Espin, 2007), a set of implicit psychometric standards can be defined for evaluating GOMs and other measures of child progress. In particular, these standards assess the degree to which measures, when used with a particular sample, yield individual scores and aggregate group distributions that are useful; that is, individual measures produce scores with distributions that are relatively normal and with few outliers or artificial minimum and maximum scores (as in floor and ceiling effects).

Based on a broad reading of the GOM and curriculum-based measurement literature (Espin, McMaster, Rose, & Wayman, 2012), we have identified six sample characteristics that can serve as standards for identifying GOMs that produce meaningful data at any particular grade level. These standards operationalize the desirable characteristics for GOMs and provide an empirical basis for selecting one measure or set of measures in any domain and grade level. Possible psychometric standards include the following:

- *Standard deviation (SD) < 50% of sample mean:* With tight distributions around sample means and fewer outliers, measures have greater capacity to discriminate differences in higher- and lower-performing students.
- *Skew < absolute value of 1:* Skew evaluates discrepancies from normality by determining the asymmetry of a distribution about its mean (Hopkins & Weeks, 1990); absence of asymmetry in score distributions is consistent with minimal floor or ceiling effects in individual measures.
- *Kurtosis < absolute value of 1:* DeCarlo (1997) reports that "kurtosis represents a movement of mass that does not affect the variance" (p. 294). Complementing skew, measures with acceptable levels of kurtosis are more likely to yield useful, well-distributed measures across individuals.
- *Test-retest reliability > .60:* GOMs must demonstrate short-term temporal stability. Test-retest reliability is a typical method to determine stability. GOMs should demonstrate a test-retest reliability of at least .60, with

strong preference for test-retest coefficients ≥.7, with coefficients .9 considered excellent (Litwin, 2002).

Sensitivity to growth: While studies have documented the test-retest reliability of GOMs in reading (e.g., Goffreda & DiPerna, 2010), such temporal stability is a double-edged sword: stability over short intervals of time provide evidence of test reliability, but stability over longer periods of time and in the context of effective intervention should not be expected. Intervention, particularly high-intensity intervention for lower-performing students, would be expected to change rank ordering of students and, in turn, reduce estimates of correlation across time. As a result, in addition to short-term temporal stability, evidence of sensitivity to growth for individual students is essential. In early cross-sectional research on any measure, growth may be inferred by relating test performance to age or grade (i.e., positive relations indicate scores increasing across time). As research continues, however, growth should be modeled directly through two or more assessments nested within children (e.g., repeated measures analysis of variance, hierarchical linear modeling).

Under 20% of target sample scoring 0: Measures that are too difficult for individuals or groups at earlier levels of skill development produce high rates of individuals scoring zero items correct. Such scores distort descriptions of performance and reduce the utility of assessment. More than a modest rate of zeros in a sample indicates floor effects or a lack of sensitivity in the measure's ability to assess performance (Carta, Greenwood, Walker, & Buzhardt, 2010). With few zeros, a tool is more likely sensitive to growth and free of influential floor effects.

These characteristics can provide crucial standards to select a set of measures that can be used in research and practice coherently over time and can help address tricky dilemmas in "age 3 to Grade 3" interventions: how to assess the growth and effects of an intervention across time as the topography of academic responding changes and how to assess reading among students who are developing essential skills but unable to decode or perform other early reading tasks. With these standards, researchers can logically identify measures that would be appropriate samples of early literacy within and across grades, all related to important long-term outcomes. As this selection process continues, attention can turn to more careful analysis of sources of variance in promising individual measures—for instance, studies of differential performance of items and scale scores across groups or, when appropriate, item- and person-fit estimates more traditionally associated with item response theory (Wilson, 2005). Measures meeting these preliminary criteria might also be assessed for sensitivity and specificity in identifying individual children who, on the basis of external criteria, might be candidates for additional intervention (Burns, Haegele, & Peterson-Brown, 2014).

Research reported here was conducted as part of a larger effort to produce a system of early literacy and beginning reading assessment for preschool and early elementary school by identifying an aligned suite of procedures for assessment from early literacy through beginning reading (from age 3 to Grade 2) for children who are developing age-typical literacy and reading skills (Wallace et al., 2007). We collected a variety of measures, including GOMs and standardized tests, using a sample of typically developing children ranging from 3 years old to Grade 2. We applied our six standards to each measure within age groups and across ages to identify measures that demonstrate adequate quality within a particular age group and that predict performance at a later age. Specific questions that we addressed were as follows:

Question 1: What are the basic sample characteristics of early literacy and beginning reading measures within grades?

Question 2: To what extent do these measures demonstrate characteristics of GOMs across grades, including balanced SD and mean, normal skew and kurtosis, sensitivity to growth, temporal stability, and low percentage of zero scores?

Question 3: To what extent do these measures relate to one another?

We then summarized results to descriptively illustrate whether student performance across these measures suggests a sequence of development and evidence to support a model of seamless assessment of early literacy and reading over time.

Methods

Participants and Settings

Participants were drawn from an initial pool of 340 children between 3 years (36 months) and 8 years 10 months (106 months) of age. Sampling included classrooms that served typically developing students and those receiving special education; however, small numbers of special education students (n = 24), coupled with uneven distribution of these students across grades and a lack of detailed information regarding disability status, led to our dropping these children from current analyses. Beyond this, we did not limit participation on gender, SES, or ethnicity. However, we excluded nonnative English-speaking students, due to evidence of differential effects on language and literacy measures for this population (Lindsey, Manis, & Bailey, 2003). Results here are for a total sample of 316 children.

The sample was sorted into grade groups for analyses. For children enrolled in elementary schools, current grade assignment (K, 1, or 2) was recorded. For children in preschool settings, age in relation to statewide eligibility for K entry was used: Children who were age eligible to enroll in K in the following academic year were assigned PK4, and children eligible for K entry two academic years hence were assigned PK3. Table 1 presents demographic information by grade. The sample was primarily White/Caucasian (mean percentage across grades, 72%) and represented students who came from homes with parents with at least a high school diploma (mean percentage across grades, 92%). Of the total sample, an average of 11% of students received free and reduced-price lunch. These demographic data were collected by direct survey request to parents of preschool participants and from school district administrative data for K and Grade 1 and 2 participants.

TABLE 1
Demographic Variables for Entire Sample by Sample, Sex, Free/Reduced-Price Lunch Status, Ethnicity, Parent Education by Grade

Age/Grade (*n*; Mean, Range)			
36 mo/PK3 (*n* = 44; *M* = 42.7 mo, 35–59)			
Males	23 (52.3)	F-RPL	1 (2.3)
Ethnicity		Parent level of education	
White/Caucasian	31 (72)	Some high school	0 (0)
African American	3 (7)	High school diploma	3 (7)
Asian American	1 (2)	Some college	2 (4)
Latino	4 (8.5)	Associate or bachelor	20 (45)
Other	1 (2)	Graduate degree	13 (30)
Did not respond	4 (8.5)	Did not respond	6 (14)
48 mo/PK4 (*n* = 69; *M* = 54.47 mo, 42–62)			
Males	33 (47.8)	F-RPL	11 (15.9)
Ethnicity		Parent level of education	
White/Caucasian	52 (76)	Some high school	1 (1)
African American	6 (9)	High school diploma	12 (17)
Asian American	3 (4)	Some college	6 (9)
Latino	3 (4)	Associate or bachelor	32 (46)
Other	1 (1)	Graduate degree	13 (19)
Did not respond	4 (6)	Did not respond	5 (8)
Kindergarten (*n* = 69; *M* = 70.48 mo, 61–78)			
Males	35 (50.7)	F-RPL	7 (10.1)
Ethnicity		Parent level of education	
White/Caucasian	51 (74)	Some high school	0 (0)
African American	1 (1)	High school diploma	9 (13)
Asian American	4 (6)	Some college	8 (12)
Latino	13 (19)	Associate or bachelor	34 (49)
Other	0 (0)	Graduate degree	12 (17)
Did not respond	0 (0)	Did not respond	6 (9)
Grade 1 (*n* = 71; *M* = 81.93 mo, 74–90)			
Males	33 (46.5)	F-RPL	13 (18.3)
Ethnicity		Parent level of education	
White/Caucasian	50 (70)	Some high school	1 (1)
African American	5 (7)	High school diploma	11 (15)
Asian American	2 (3)	Some college	3 (4)
Latino	14 (20)	Associate or bachelor	35 (49)
Other	0 (0)	Graduate degree	19 (28)
Did not respond	0 (0)	Did not respond	2 (3)
Grade 2 (*n* = 63; *M* = 94.98 mo, 86–103)			
Males	38 (60.3)	F-RPL	6 (9.5)
Ethnicity		Parent level of education	
White/Caucasian	43 (68)	Some high school	0 (0)
African American	4 (6)	High school diploma	4 (6)
Asian American	5 (8)	Some college	10 (16)
Latino	11 (18)	Associate or bachelor	26 (41)
Other	0 (0)	Graduate degree	22 (36)
Did not respond	0 (0)	Did not respond	1 (1)

Note. Values presented as *n* (%) unless noted otherwise. F-RPL = free/reduced-price lunch; PK = prekindergarten (for PK3 and PK4, see Participants and Settings).

Measures

The early literacy and beginning reading assessment protocol included 11 measures (10 developed as GOMs and 1 standardized). We chose measures to assess language development, phonological awareness, letter identification and letter-sound correspondence, and reading. We used known measures representing one or more aspects of early literacy and demonstrating initial evidence of reliability and validity (e.g., NELP, 2009; Wayman et al., 2007).

Language development. IGDI Picture Naming (PN; Early Childhood Research Institute on Measuring Growth and Development [ECRI-MGD], 1998) is an individually administered 1-minute measure of expressive language for 3- to 6-year-olds. With 120 randomly ordered photos and drawings of everyday objects labeled in the lexicon of typical 5-year-olds, the administrator (a) described and demonstrated the task to the child, (b) asked the child to "name the pictures as quickly as you can," and (c) showed the cards sequentially for 1 minute (for details, see http://www.myigdis.com). PN shows moderate to high alternate-form reliability and concurrent validity with established norm-referenced measures of preschool skills (ECRI-MGD, 2004; Priest, Davis, McConnell, McEvoy, & Shinn, 1999). Missall and McConnell (ECRI-MGD, 2004) found it to be sensitive to growing expressive language skills. The score is the number of pictures named correctly in 1 minute.

Alphabetic principle. DIBELS Letter Naming Fluency (LNF; Good et al., 2002) is an individually administered 1-minute task. The examiner showed the student upper- and lowercase letters, randomly ordered and arranged in rows, and then asked the child to name as many letters as possible. The student's final score is the number of letters named correctly (Assessment Committee Analysis of Reading Assessment Measures [ACARAM], 2002; Good et al., 2002). LNF demonstrated high alternate-form reliability

(ACARAM, 2002; Good et al., 2002) and high concurrent validity relations with the Woodcock-Johnson Psycho-Educational Battery Total Reading Cluster, Stanford Diagnostic Reading Test at Diagnostic Reading Assessment, and the Test of Early Reading Ability (ACARAM, 2002; Good et al., 2002; Rouse & Fantuzzo, 2006).

Minneapolis Kindergarten Beginning Assessment Letter Sounds, a standardized format developed locally and based on other reported measures (Research Evaluation and Assessment, 2004), was used to assess letter-sound correspondence. The administrator presented the child with a sheet containing 100 randomly ordered upper- and lowercase letters. The administrator stated, "I'm going to show you some letters and I want you to make the sound the letter makes." The child's score is the number of correct responses in 1 minute. Analysis of this measure during its development yielded 4-week test-retest reliability at .89 and evidence of moderately high internal consistency.

Phonological awareness. IGDI Rhyming is an individually administered 2-minute task in which children identify two rhyming words (a target stimulus and one of three distractors; ECRI-MGD, 2004). The test includes about 60 randomly ordered stimulus cards. Each card has four pictures: a target in the top row center and three below. The assessor labeled each picture and asked the child to identify the one in the bottom row that rhymed with the top one. Correct responses measure rhyme awareness (see http://www.myigdis.com). Rhyming scores are stable for 3 weeks and correlate with other measures of language and literacy development, including the Peabody Picture Vocabulary Test–Third Edition (PPVT-3; r = .56–.62), Concepts About Print (r = .54–.64; Clay, 1985), and the Test of Phonological Awareness (r = .44–62; Torgesen & Bryant, 1994; see also, ECRI-MGD, 2004; McConnell, McEvoy, & Priest, 2002; Missall 2002; Priest, Silberglitt, Hall, & Estrem, 2000).

IGDI Alliteration is an individually administered 2-minute task in which children identify two words (a target and one of three distractors) that share an initial sound (ECRI-MGD, 2004). The test includes approximately 60 randomly ordered stimulus cards, each containing four pictures: a target in the top row and three choices below. The assessor labels each picture and asks the child to identify the bottom one that begins with the same sound as the top one. Correct responses record a measure of alliteration. Alliteration scores are moderately stable over 3 weeks (ECRI-MGD, 2004) and correlate with other measures of language and early literacy, including the PPVT-3, Test of Phonological Awareness, and Concepts About Print.

IGDI Sound Blending (SB) is an individually administered 2-minute task in which children produce words after being presented with stimulus items segmented at the level of compound words (e.g., cow/boy), syllable (e.g., ta/ble), or phoneme (e.g., s/a/t). The student views about 30 segmented words. The number of correct blends produced served as the score. Limited analyses of psychometric properties of this task took place during initial development and testing, with analyses based on samples of 30 to 90 children. Test-retest over 2 weeks demonstrated moderate stability (.73). The measure also demonstrated moderate relations to individually administered measures of oral language development, including PPVT-3 and Concepts About Print.

DIBELS Phoneme Segmentation Fluency (PSF) is an early literacy measure of phonemic awareness measuring a child's ability to segment simple words into single phonemes. The examiner speaks words with three to four phonemes and asks the student to state the individual phonemes in each word (e.g., the examiner says the word *sat*; the student says "/s/ /a/ /t/" to receive three possible points). The number of correct phonemes in 1 minute is the score. PSF alternate-form reliability ranges from .60 to .88 (ACARAM, 2002; Good et al., 2002), with moderate to strong relations to concurrent validity measures, including PSF in later grades, the Metropolitan Readiness Test, and the Stanford Diagnostic Reading Test (Good et al., 2002).

Beginning reading. DIBELS Nonsense Word Fluency (NWF) is an early literacy test of alphabetic principle and blending. To administer the test, an examiner shows the student a paper with randomly ordered vowel/consonant and consonant/vowel/consonant nonsense words. The administrator asks the student to vocalize individual sounds of each letter or read the whole nonsense word. For example, if the stimulus word is "rav," the student could say /r/ /a/ /v/ or say the word "rav" to earn a score of 3. The score is the total number of correct letter sounds in 1 minute (Good et al., 2002). Alternate-form reliability has been tested through single- and multiprobe methods. Reliability coefficients range from .92 to .98 (ACARAM, 2002; Good et al., 2002). Predictive and concurrent validity has also been shown. NWF correlates with DIBELS PSF (.59; Good et al., 2002), and predictive validity indicates that NWF correlates with the Woodcock-Johnson Psycho-Educational Battery Total Reading Cluster score (.66) and Curriculum-Based Measurement–Oral Reading (CBM-OR; .82; ACARAM, 2002; Good et al., 2002). Concurrent validity coefficients with the Test of Early Reading Ability and Diagnostic Reading Assessment range from .35 to .62 (Rouse & Fantuzzo, 2006).

Dolch Word List includes 220 high-frequency words published by Edward Dolch in 1948. For this study, the administrator showed the first 50 words of the list on a sheet of paper to students and asked them to read the words as quickly and correctly as they could. The administrator told students to go on to the next word if they did not know a word. Words read correctly were recorded for a 1 minute. We located no empirical evaluation of reliability and validity for the Dolch Word List but included it due to the frequency with which investigators use the list (e.g., Bliss, Skinner, & Adams, 2006; Reifman, Pascarella, & Larson, 1981).

CBM-OR is an assessment of reading accuracy and fluency with connected text. Students receive a series of passages taken from standardized publications or grade-level curricula. Passages are calibrated by grade level. Student performance is measured as students read the passages aloud for 1 minute. Omitted or substituted words and hesitations >3 seconds are counted as errors. After students read the passages, the median score of words read correctly becomes the oral reading fluency rate (Deno et al., 1982). We obtained the passages with permission from Vanderbilt University courtesy of Dr. Lynn Fuchs (see Wallace et al., 2007). Test-retest reliability ranges from .92 to .97 and alternate-form reliability from .89 to .94 (Deno et al., 1982; Tindal, Marston, & Deno, 1983). Eight separate studies reporting coefficients from .52 to .91 established criterion validity (Good & Jefferson, 1998; Marston & Magnusson, 1988; Ysseldyke et al., 1983).

Standardized measure. Woodcock-Johnson Test of Achievement–III (Woodcock, McGrew, & Mathers, 2002) is an individually administered norm-referenced test to assess reading, oral language, mathematics, written language, and academic knowledge. We used only the Letter-Word Identification (LWI) subtest, a test of basic literacy skills involving symbolic learning and identifying isolated letters and words. The child identifies letters that are in large type and reads the words correctly. Items are set in difficulty order, with the easiest first and the most difficult last. Testing stops when the student scores zero on six successive items (Woodcock et al., 2002).

Procedures

After receiving parental consent, we tested students on some or all of the 11 measures, based on likely performance by grade. Preschoolers completed all measures except CBM-OR. First and second graders did not complete IGDI Alliteration or Letter Naming (LN), due to evidence of significant ceiling effects in prior pilot testing. Each test was administered in two individual sessions 12 weeks apart. Trained graduate research assistants assessed the participants in or immediately outside the classroom.

Design and Data Analysis

To answer Question 1—*What are the basic sample characteristics of early literacy and beginning reading measures within grades?*—we examined mean, median, range, skew, kurtosis and normality of each measure within each grade group.

For Question 2—*To what extent do these measures demonstrate characteristics of GOMs across grades, including balanced SD and mean, normal skew and kurtosis, sensitivity to growth, temporal stability, and low percentage of zero scores?*—we evaluated each measure collected for each grade against our proposed six psychometric standards to determine use as an acceptable GOM. First, we calculated a ratio of the SD to sample mean by grade, and we identified those instances where the SD was <50% of the sample mean. Next, we evaluated measures by grade where skew and kurtosis were each less than an absolute value of 1. We also noted those measures where <20% of a sample failed to obtain a single correct response. We modeled stability over time by calculating test-retest reliability. Sensitivity to growth was determined by significant within-subject differences from Time 1 to Time 2 as well as slopes >1. Effect sizes for differences between Time 1 and Time 2 are reported, with p values reported for testing across-time differences and nonzero slope. To summarize analyses, we calculated the number of psychometric standards met by grade group, with a heuristic criterion that defined measures as "acceptable" for a particular grade if they met four of six psychometric standards.

To answer the third question—*To what extent do these measures relate to one another?*—we examined concurrent correlations across measures collected at Time 1. To evaluate our final point—*Does student performance across these measures suggest a sequence of development and a model for a "seamless" system of assessment over time?*—we examined mean, median, range, skew, kurtosis, and normality of each measure across grade groups to ascertain the extent to which measures aligned seamlessly over time.

Results

Sample and Psychometric Characteristics

Table 2 presents descriptive statistics, including means and SDs for all measures. Table 3 presents results of comparisons to the six GOM standards. Assumptions regarding normality, outliers, linearity, and homogeneity of variance were not violated for any measure.

Language development. In general, distributions were relatively normal, with acceptable skew and kurtosis. PN slopes differed significantly from zero (with significant difference in performance between Time 1 and Time 2) for PK4, K, and Grade 1. Reliability estimates were low-moderate to moderate. PN met five of six standards for PK3, PK4, K, and Grade 1 and four of six standards for Grade 2.

Alphabetic principle. LNF, Letter Sound Fluency (LSF), and LWI measured alphabetical principle. LNF distributions varied somewhat from desired characteristics, with SDs (particularly for PK3 and PK4) ranging from 0.74 to 1.94 of the sample mean and with skew and kurtosis substantially above standards for two of three assessed grades. Change over time was detected at all three grades. Test-retest reliability was in the range of high-moderate to high. LNF met all six quality standards for K, five of six for PK4, and two of six for PK3.

TABLE 2
Early Literacy Measure by Grade Descriptive Statistics

	PN			LNF			LS			WJ-LWI			Rhyming			Alliteration		
Grade	*n*	*M*	*SD*	*n*	*M*	*SD*	*n*	*M*	*SD*	*n*	*M*	*SD*	*n*	*M*	*SD*	*n*	*M*	*SD*
PK3	44	16.88	7.13	44	9.12	11.91	44	1.95	3.06	44	6.24	4.24	41	3.17	3.75	41	1.73	2.19
PK4	69	22.74	7.69	69	20.86	15.31	69	5.40	7.54	69	10.96	5.74	69	6.63	5.94	69	4.10	4.96
K	69	27.54	7.84	16	9.14	17.72	69	29.17	13.62	69	19.88	7.04	69	14.43	5.06	69	13.07	4.89
1	71	28.38	7.06	2	55.00	8.49	71	43.86	12.52	71	36.63	8.95	71	16.37	6.68	2	10.50	3.54
2	63	34.19	7.26	1	NA	NA	63	44.37	13.61	63	47.44	9.66	63	20.89	5.59	1	NA	NA

	SB			PSF			NWF			Dolch Word List			CBM-OR		
PK3	44	5.46	5.12	44	1.15	5.19	44	2.33	12.95	44	1.21	7.53	0	NA	NA
PK4	69	10.97	8.23	48	1.92	4.15	48	3.47	7.91	49	1.04	5.81	0	NA	NA
K	69	23.67	7.18	69	30.17	13.95	69	28.93	29.79	16	0.14	0.54	0	NA	NA
1	71	26.65	7.26	69	46.01	11.71	69	69.49	35.67	71	39.01	28.87	69	63.12	44.34
2	63	35.25	9.13	62	43.47	13.27	62	105.31	41.23	68	74.57	24.85	62	114.05	37.56

Note. CBM-OR = Curriculum-Based Measurement–Oral Reading; K = kindergarten; LNF = Letter Naming Fluency; LS = Letter Sounds; NA = not applicable; NWF = Nonsense Word Fluency; PN = Picture Naming; PK = prekindergarten (for PK3 and PK4, see Participants and Settings); PSF = Phoneme Segmentation Fluency; SB = Sound Blending; WJ-LWI = Woodcock-Johnson Letter-Word Identification Subtest.

TABLE 3
Analysis of General Outcome Measure Characteristics by Measure: Time 1

Descriptive	PK3	PK4	Kindergarten	Grade 1	Grade 2
			Picture Naming		
M/*SD*	**0.42**	**0.34**	**0.28**	**0.25**	**0.21**
Effect size (slope)	0.06 (0.14)	**0.15**** **(0.33)****	**0.18***** **(0.34)*****	**0.13**** **(0.26)****	0.05 (0.15)
% of zeros	**2**	**1**	**0**	**0**	**0**
Maximum[a]	30	38	38	30	37
Skewness	**–0.03**	**–0.25**	**–0.41**	**0.12**	**0.12**
Kurtosis	**–0.74**	**–0.17**	**0.21**	**–0.65**	**0.07**
Reliability	**0.64****	0.55**	0.51**	0.44**	0.55**
			Letter Naming Fluency		
M/*SD*	1.31	0.74	**0.35**	NA	NA
Effect size (slope)	**0.13*** **(0.29)***	**0.07*** **(0.30)***	**0.06*** **(0.27)***	NA	NA
% of zeros	25	**10**	**0**	NA	NA
Maximum[a]	33	58	72	NA	NA
Skewness	2.29	**0.39**	**–0.35**	NA	NA
Kurtosis	7.37	**–0.57**	**–2.05**	NA	NA
Reliability	**0.87****	**0.79****	**0.76****	NA	NA
			Letter Sounds		
M/*SD*	1.57	1.40	**0.47**	**0.29**	**0.31**
Effect size (slope)	0.00 (0.00)	**0.22***** **(0.29)*****	**0.12**** **(0.38)****	0.01 (0.12)	**0.08*** **(0.31)***
% of zeros	55	34	**1**	**0**	**0**
Maximum[a]	11	43	69	73	90
Skewness	1.80	2.57	**0.58**	**–0.21**	**0.66**
Kurtosis	2.62	8.95	**0.67**	**–0.11**	1.57
Reliability	**0.72****	0.56**	**0.66****	**0.65****	**0.66****

(continued)

TABLE 3. (CONTINUED)

Descriptive	PK3	PK4	Kindergarten	Grade 1	Grade 2
			Woodcock-Johnson III Letter-Word Identification Subtest		
M/*SD*	0.68	**0.52**	**0.35**	**0.24**	**0.20**
Effect size (slope)	0.07 (0.21)	0.01 (0.03)	**0.39***** **(0.29)*****	**0.58***** **(0.43)*****	0.02 (0.07)
% of zeros	**9**	**2**	**0**	**0**	**0**
Maximum[a]	19	40	33	40	38
Skewness	**0.66**	2.11	1.75	**0.59**	**–0.01**
Kurtosis	**0.57**	10.00	3.12	**0.27**	**–0.58**
Reliability	0.24	**0.72****	**0.88****	**0.90****	**0.82****
			Rhyming		
M/*SD*	1.18	0.90	**0.35**	**0.41**	**0.27**
Effect size (slope)	**0.18**** **(0.17)****	**0.10*** **(0.17)****	**0.32***** **(0.32)*****	**0.36***** **(0.29)*****	**0.22***** **(0.27)*****
% of zeros	45	33	**3**	**4**	**1**
Maximum[a]	13	22	27	32	33
Skewness	**0.88**	**0.46**	**–0.30**	**–0.12**	**–0.58**
Kurtosis	**–0.30**	**–0.67**	1.48	**0.67**	2.15
Reliability	**0.73****	**0.68****	**0.69****	**0.78****	0.57**
			Alliteration		
M/*SD*	1.27	1.21	**0.37**	NA	NA
Effect size (slope)	0.06 (0.10)	**0.08*** **(0.14)***	**0.36***** **(0.29)*****	NA	NA
% of zeros	50	43	1	NA	NA
Maximum[a]	7	19	15	NA	NA
Skewness	**0.96**	1.25	**0.38**	NA	NA
Kurtosis	**–0.19**	**0.93**	1.67	NA	NA
Reliability	0.22	**0.71****	**0.74****	NA	NA
			Sound Blending		
M/*SD*	0.94	0.75	**0.30**	**0.27**	**0.26**
Effect size (slope)	0.02 (.04)	**0.19***** **(0.38)*****	**0.40***** **(0.59)*****	**0.43***** **(0.66)*****	0.01 (0.10)
% of zeros	22	**8**	**0**	**0**	**0**
Maximum[a]	3	20	32	30	66
Skewness	**0.73**	**0.59**	**–0.51**	**–0.18**	3.87
Kurtosis	**–0.27**	**–0.08**	**–0.12**	**–0.61**	21.06
Reliability	**0.73****	**0.63****	0.51**	0.28*	–0.03
			Phoneme Segmentation Fluency		
M/*SD*	4.51	2.16	**0.46**	**0.25**	**0.31**
Effect size (slope)	0.00 (–0.04)	**0.15**** **(0.23)***	**0.07*** **(0.38)***	0.01 (–0.11)	0.10* (–0.43)*
% of zeros	75	49	**3**	**0**	**0**
Maximum[a]	32	16	66	74	66
Skewness	5.84	2.46	**0.05**	**–0.18**	**–0.40**
Kurtosis	35.27	5.40	**0.28**	**–0.18**	**–0.53**
Reliability	0.21	0.21	0.42**	0.26**	0.38**
			Nonsense Word Fluency		
M/*SD*	5.50	2.28	1.03	0.51	**0.39**
Effect size (slope)	0.02 (–0.08)	**0.18**** **(0.25)****	**0.10**** **(0.51)****	**0.20***** **(0.91)*****	0.00 (0.15)
% of zeros	75	50	**8**	**0**	**1**
Maximum[a]	81	40	148	179	237
Skewness	6.21	3.04	2.30	1.10	**0.22**
Kurtosis	38.67	10.21	**6.14**	**1.17**	**0.67**
Reliability	**0.91****	**0.68****	**0.87****	**0.85****	**0.69****

(continued)

TABLE 3. (CONTINUED)

Descriptive	PK3	PK4	Kindergarten	Grade 1	Grade 2
			Dolch Word List		
M/*SD*	6.22	5.59	3.86	0.74	**0.33**
Effect size (slope)	0.02 (–0.02)	0.07 (0.03)	0.01 (0.11)	**0.46***** **(1.32)*****	0.01 (0.15)
% of zeros	86	62	87	**0**	**0**
Maximum[a]	47	41	92	95	110
Skewness	6.25	6.91	3.74	**0.64**	–1.06
Kurtosis	39.00	48.44	14.00	**–0.93**	**0.82**
Reliability	**0.95****	0.29*	**0.81****	**0.85****	**0.70****
			Curriculum-Based Measurement–Oral Reading		
M/*SD*	NA	NA	NA	0.70	**0.33**
Effect size (slope)	NA	NA	NA	**0.70***** **(1.88)*****	**0.12**** **(0.79)****
% of zeros	NA	NA	NA	**0**	**0**
Maximum[a]	NA	NA	NA	194	178
Skewness	NA	NA	NA	**0.92**	**–0.08**
Kurtosis	NA	NA	NA	**0.22**	**0.09**
Reliability	NA	NA	NA	**0.95****	**0.85****

Note. Bold indicates those values meeting the general outcome measure criteria. Time 1 and Time 2 scores were used to evaluate the effect size (partial η^2) of the slope and reliability. Differences between the slope and zero as well as differences between Time 1 and Time 2 were used to examine probability value, as indicated by asterisks. Maximum over time was evaluated with repeated measures analysis of variance, within-subjects effect. NA = not applicable; PK = prekindergarten (for PK3 and PK4, see Participants and Settings).

[a]Minimum for all measures included observed zero scores.

$^{*}p < .05$. $^{**}p < .01$. $^{***}p < .001$.

Measures of distributions for LSF also showed variation from desired standards. Change over time was noted for three grades (PK4, K, and Grade 2); temporal stability met standards for four of five grades (all but PK3). LSF met all six standards for K and five standards for Grades 1 and 2. In PK4, LSF met only the standard for sensitivity to growth; in PK3, this measure met only the criterion for reliability.

LWI met all six standards for Grade 1, five standards for Grade 2, four standards for K, and three of six for PK3 and PK4.

Phonological awareness

Rhyming. For Rhyming, distributions were within standards for K and Grades 1 and 2, with SDs varying from 0.27 to 1.18 of the sample mean across all five grades. Sensitivity to growth was detected at all grades, and skew and kurtosis estimates ranged from |0.12| to |2.15|; skew met our a priori standard (<|1|) for all five grades, and kurtosis met this standard in three. Test-retest reliability was in the high-moderate range for all grades, although this estimate did not meet a priori standard for Grade 2. A large proportion of PK3 and PK4 children scored zero. Rhyming met all six standards for Grade 1; five for K; and four for Grade 2, PK3, and PK4. It demonstrated SD values >50% of the mean and a high percentage of zero scores for PK3 and PK4 while obtaining lower reliability (.57) and a large kurtosis for Grade 2.

Alliteration. Only PK3, PK4, and K students took the Alliteration test. In general, SDs exceeded standards when compared with means for PK3 and PK4 participants. Distributions had significant skew for PK4 students and met standard for PK3 and K. Sensitivity to growth improved by grade, and temporal reliability was low to moderate. The measure met five standards in K (with only kurtosis outside standard levels), three standards in PK4, and two standards in PK3.

Sound Blending. SB met half or more a priori standards for all grades. All distributions met standards for relations between mean and SD, and all but K students met standards for skew and kurtosis. Other than PK3 students, where 22% of the sample scored zero, almost all students scored above zero. Slope and differences between Times 1 and 2 were different from zero for PK4, K, and Grade 1 students. Five standards were met for K, Grade 1, and PK4.

Phoneme Segmentation Fluency. For PSF, the relative size of SDs varied across grades, with ratios meeting the a priori standard in only K and Grades 1 and 2. Skew and kurtosis varied similarly, with values at K and Grades 1 and 2 at or near zero. Sensitivity to growth was detected at PK4 and K. Growth was also detected at Grade 2 to be significantly different from zero, however in a negative direction (slope, –0.43). PSF met five of six standards for K and four standards for Grades 1 and 2, with data at Grades 1 and 2 not meeting standards for sensitivity to growth or reliability. PK4 met one standard, sensitivity to growth, while PK3 met none of the six.

Beginning reading

Nonsense Word Fluency. Characteristics of sample distributions for NWF varied across grades. Skewness and kurtosis exceeded |2.3| for PK3, PK4, and K and were close to or below |1| for Grades 1 and 2. Sensitivity to growth was detected only during PK4, K, and Grade 1. Test-retest reliability was moderate to high across all grades. Half or more of PK3 and PK4 students scored zero on this measure, but much smaller percentages of zero scores were found in K and Grades 1 and 2. NWF met five standards in Grade 2, four standards for K and Grade 1, two standards for PK4, and one standard for PK3.

Dolch Word List. SD ratios to sample means ranged from 3.86 to 6.22, and skewness and kurtosis estimates exceeded |3.74| for PK3, PK4, and K students. SD-to-mean ratios were .33 to .74 and skewness or kurtosis estimates were close to or below |1| for Grades 1 and 2. Sensitivity to growth was only detected in Grade 1; test-retest reliability was moderate to high across all grades except PK4, and 62% to 87% of PK3, PK4, and K students scored zero. Dolch met five standards for Grade 1, four standards for Grade 2, one standard for PK3 and K, and none for PK4 students.

Curriculum-Based Measure–Oral Reading. Only Grades 1 and 2 took the CBM-OR test. Relations between means and SDs met standard; kurtosis was close to zero in both grades; and skew met standards for both grades. Test-retest reliability estimates exceeded .85, and no students had a zero score. CBM-OR was sensitive to growth over time for both grades. CBM-OR for Grade 2 met all six standards, and Grade 1 met five standards.

Summary across measures for each grade. Table 4 summarizes GOM standards met for each measure at each grade. We noted some variability across measures and grade groups, with evidence of a developmental trend: When variability in number of standards existed, the number of standards met generally increased with grade. Two measures, PN and Rhyming, met four or more standards across all grades. Several measures functioned well in one or more, but not all, groups, and each measure failed to meet at least one standard in one or more grade groups. Only PN met five standards for PK3; three measures (PN, SB, and LNF) met five standards for PK4. In K, five measures met five standards, and two (LSF and LNF) met all six. All measures met at least four standards for Grade 1, with two meeting all six (LWI and Rhyming). Finally, four measures met at least five standards for Grade 2.

TABLE 4
Number of General Outcome Measure Standards Met by Measure and Grade Group

	Grade				
	PK3	PK4	K	1	2
Language: Picture Naming	5	5	5	5	4
Alphabetic principle					
Letter Naming Fluency	3	5	6	NA	NA
Letter Sounds	1	1	6	5	5
WJ-LWI	3	3	4	6	5
Phonologic awareness					
Rhyming	4	4	5	6	4
Alliteration	2	3	5	NA	NA
Sound Blending	3	5	5	5	2
Phoneme Segmentation Fluency	0	1	5	4	4
Beginning reading					
Nonsense Word Fluency	1	2	4	4	5
Dolch Word List	1	0	1	5	4
CBM-OR	NA	NA	NA	5	6

Note. CBM-OR = Curriculum-Based Measurement–Oral Reading; NA = not applicable (measure not collected); PK = prekindergarten (for PK3 and PK4, see Participants and Settings); WJ-LWI = Woodcock-Johnson Letter-Word Identification Subtest.

Relations Among Measures

To examine relations among measures, we computed Pearson correlations during Time 1 (see Table 5).[3] In general across 53 unique bivariate coefficients, relations were positive and moderate to strong, with one negative coefficient, six correlations <.40, and 36 correlations (or 68% of all coefficients) ≥.60. Relations were generally strong for measures from common elements of early literacy, with coefficients ≥.82 for LN and sounds, .70 to .81 for early childhood phonological awareness measures, and .79 to .87 for reading sight or nonsense words in isolation (Dolch and NWF) as well as reading connected text (CBM-OR). Correlations across these common elements were more variable; notably, PN (putatively, a measure of vocabulary and language development) showed moderate to moderate-strong relations across all other measures except CBM-OR.

Discussion

This study's aim was to assess the extent to which each of 10 measures of early literacy met standards associated with GOMs to be used across grades, to identify measures for preschool through early elementary school to assess children's development of early literacy. The broad intent is for findings that will guide future research on assessment and intervention evaluation and modification. Using a descriptive approach, we evaluated measures at each grade from PK3 to Grade 2 against six standards for GOMs: SDs <50% of sample means, skew and kurtosis less than absolute values of 1.0, sensitivity to growth over 12 weeks, short-term test-retest reliability exceeding .60, and <20% zero scores.

TABLE 5
Correlation Matrix Between and Among Measures at Time 1 Across Grade Levels: PK3 to Grade 2

	LNF	LS	WJ-LWI Raw	Rhyming	Alliteration	SB	PSF	NWF	Dolch	CBM-OR
PN	.55**	.55**	.57**	.67**	.58**	.66**	.53**	.49**	.49**	.33**
LNF		.85**	.82**	.71**	.76**	.79**	.71**	.54**	.51**	NA
LS			.80**	.72**	.82**	.79**	.74**	.48**	.65**	.14
WJ-LWI raw				.75**	.77**	.77**	.72**	.89**	.87**	.84**
Rhyming					.81**	.74**	.70**	.65**	.61**	.37**
Alliteration						.78**	.75**	.61**	.39**	NA
SB							.78**	.69**	.64**	.35*
PSF								.65**	.57**	–.16
NWF									.89**	.79**
Dolch										.87**

Note. CBM-OR = Curriculum-Based Measurement–Oral Reading; LNF = Letter Naming Fluency; LS = Letter Sounds; NA = not applicable; NWF = Nonsense Word Fluency; PN = Picture Naming; PK = prekindergarten (for PK3, see Participants and Settings); PSF = Phoneme Segmentation Fluency; SB = Sound Blending; WJ-LWI = Woodcock-Johnson Letter-Word Identification Subtest.
*$p < .05$. **$p < .01$. ***$p < .001$.

Results indicated variation across measures and grades in meeting heuristic standards of acceptability for GOMs within and across grades. Across grades and potential GOMs, results indicate that three measures (PN, Rhyming, and CBM-OR) met at least four standards across all grades tested, with PN and Rhyming meeting at least four of six GOM standards across all grades from preschool through early elementary school. Several other measures showed some promise for cross-grade application, particularly phonological awareness measures (i.e., Rhyming and SB).

Results suggest that several measures might be used across three or more grades. This was particularly true for K to Grades 1 and 2: Six measures (PN, LSF, NWF, Woodcock-Johnson Test of Achievement–III, Rhyming, and PSF) met four or more standards for all three early elementary grades. Similarly, three measures met four or more standards for the transition from preschool to elementary school (i.e., PK4 to Grade 1). Results for LNF and LSF were intriguing but not definitive. Future research might evaluate the utility of these measures, individually or collectively, for continuous assessment of students from PK4 to Grade 1.

Implications for Research and Practice

These results have implications for ongoing research on GOMs, as well as early intervention to promote long-term reading proficiency. For research, when considered with analyses of longitudinal relations of preschool and early elementary measures (Missall et al., 2007) and analyses of reading measures for elementary and secondary students (Wayman et al., 2007) as well as burgeoning work evaluating long-term effects of early childhood services (e.g., Burchinal, Vandergrift, Pianta, & Mashburn, 2010; Lonigan et al., 2015), these results support efforts to develop an integrated, seamless assessment system of children's paths toward proficient reading starting in preschool. Central to such a system would be measures to mark progress within and across service delivery boundaries (e.g., preschool; preschool to K; K to Grades 1 and 2). This assessment of child progress within and across boundaries on a set of measures available to researchers and practitioners creates a stronger conceptual and empirical basis for asserting that intervention in one setting, focused on potentially different topographies of child behavior and using different curricular and instructional procedures, will produce effects on child performance in a subsequent setting and a seemingly different class of behaviors (cf. Fuchs & Deno, 1991). It also positions research findings to more rapidly transfer to implementation in practice settings through the use of common measures (Bijou, Peterson, & Ault, 1968).

At a minimum, such a system would require that (a) measures collected in an earlier setting be highly related to measures in subsequent settings and (b) measures collected in both settings be highly related to socially valid (and typically distal) general outcomes (cf. Deno, 1997; Fuchs & Deno, 1991). Results presented here suggest several measures that may meet these standards across some settings or grades. In particular, PN and Rhyming met many of the identified standards for GOMs for three or more adjacent grade groups covering preschool and early elementary (K) grades. While results indicate that some refinement or improvement of individual measures might be needed at particular grades (for instance, SB and PN for PK3), it is likely warranted to assess the function and growth trajectory of these measures in future longitudinal research.

These results offer some initial evidence and basis for considering further integration of existing theoretical models of early literacy and beginning reading. At present,

theoretical models of early literacy (NELP, 2009; Sénéchal et al., 2001; Whitehurst & Lonigan, 1998) and beginning reading (Adams, 1990; Hoover & Gough, 1990) are likely to be complementary, but little attention has been given to explicit analysis of their relations and points of connection and distinction. Such an analysis will help clarify theoretically and practically important issues, such as whether some individual measure adequately indicates development or whether one or several domain-referenced measures collectively serve as broad indicators of early literacy development.

Large-scale curricular practices and educational policy also affect this issue: If the explicit skills associated with reading are introduced formally at a particular time and if there is little interest in acquiring these skills before then, acquisition might appear to be grade-related when, in fact, it is not. Such may be the case in contemporary policy and practice in early literacy: As we adjust and change our expectations about what to teach to younger children, we may reasonably expect to see changes in "normative" performance on tasks that were previously taught in later grades. If this is true, what constitutes age-appropriate early literacy and reading development may soon change.

Finally, from an applied perspective, practitioners may benefit from understanding these relations between measures (and, perhaps, skills) within and across grades when choosing and aligning assessment and intervention procedures. With empirical information on the utility of GOMs in hand, teachers can make educated decisions about how long and to what degree each measure will be useful in data-based decision making and intervention.

Limitations

As with any investigation, several limitations of the current study should be noted. First, while grade-level samples ranged from 44 to 73 children and included some degree of variability in ethnic, economic, and parental education measures, variation in children, classrooms, and curricular offerings is necessarily limited. Such is particularly the case for students with disabilities and those who might benefit from supplemental or more intensive intervention, as provided in multitiered systems of support. These tiered intervention programs are increasingly prevalent in educational settings, and these systems frequently use measures like those studied here (Fuchs & Fuchs, 2007; Greenwood, Carta, & McConnell, 2011). Expanding investigations for relations in performance above and below standards for supplemental intervention and for children who perform significantly below grade-level expectations will yield important findings. Similarly, the sample presented in this study is limited in ethnic diversity and risk status variables, such as parent level of education and free and reduced-price lunch. Given the homogenous nature of the students and families included in this study, future research should continue to evaluate systematic approaches for evaluating a seamless collection of measures to measure the transition to reading with more diverse samples.

Measurement of educational achievement is a dynamic, fast-developing area of research. Continued evolution and development of measures like those included here (e.g., Christ & Nelson, 2014; McConnell, Wackerle-Hollman, & Bradfield, 2014) will likely improve estimates and operations across time and grades but will also necessitate new analysis of relations and applications. As measures and interventions continue to improve, research like this will need to be refined and revisited.

These results also do not answer another pressing issue in current GOM research and development: To what extent should GOMs be robust and comprehensive, versus specific and sensitive? More comprehensive measures, such as CBM-OR, likely sample a variety of child skills (e.g., letter-sound correspondence, blending, prosody, fluency, text comprehension) and are apt to show growth over longer periods. Other measures, such as Rhyming or PSF, sample a narrower band of child behaviors. While narrower measures may assess a theoretically important part of a larger skill (i.e., reading) and while a more specific assessment may lend itself to greater sensitivity within samples or across time, there could be negative consequences in relation to longer-term or more general outcomes (Deno, 1997; Paris, 2005). As measures become too specific and narrow, they may lose utility as GOMs and affect efforts to assess child progress and allocate intervention resources in ways that produce long-term benefit (Fuchs & Fuchs, 2007). This tension between generality and specificity will likely continue to vex researchers and practitioners, and it may be an important area for continued theoretical and empirical analysis.

Theoretical advances in recent decades and ongoing research articulating literacy from early childhood through elementary education lay foundations for significant gains in equity and excellence in reading and literacy outcomes. This scholarship represents ongoing efforts to build a practical, contemporary model of early literacy assessment for children in later preschool and early elementary school. This work, along with ongoing research on related and more basic topics, is likely to continue informing research, policy, and practice.

Acknowledgments

The opinions expressed in this article are those of the authors and imply no official endorsement by the funding agency. We thank Dr. Kevin McGrew for permission to reproduce portions of the Woodcock-Johnson III Letter Word Identification protocol; Dr. Lynn Fuchs for graciously sharing measures used in her group's prior research; graduate research assistants and staff from the Research Institute on Progress Monitoring and the Center for Early Education and Development at the University of Minnesota for

assisting with data collection; and coinvestigators from Research Institute on Progress Monitoring. We also thank Drs. Teri Wallace, Chris Espin, Stan Deno, and Kristen McMaster for comments on earlier versions of this report.Drs. Scott R. McConnell and Alisha Wackerle-Hollman have developed assessment tools and related resources known as Individual Growth & Development Indicators and Get it, Got it, Go! This intellectual property is the subject of technology commercialization by the University of Minnesota, and portions have been licensed to Early Learning Labs, Inc. McConnell has equity interest in Early Learning Labs, Inc., a company that may commercially benefit from the results of this research. The University of Minnesota also has equity and royalty interests in Early Learning Labs, Inc., which may in turn benefit McConnell and Wackerle-Hollman. These relationships have been reviewed and are being managed by the University of Minnesota in accordance with its conflict-of-interest policies.

Funding

This research was supported by a grant from the Research Institute on Progress Monitoring, Office of Special Education Programs, U.S. Department of Education (H324H030003).

Notes

1. For this article, we use the term *early literacy* to denote reading-related behaviors and skills developed before formal reading and reading instruction begin, typically occurring in preschool. *Beginning reading* describes skills and behaviors more akin to formal reading, typically occurring after the start of formal reading instruction in kindergarten or Grade 1.

2. While writing is included in some models of reading and literacy in elementary grades and while presumed precursors (e.g., invented spelling) have received attention in early childhood research (National Early Literacy Panel, 2009), formal assessment and intervention for writing remain somewhat rare and inconsistent in early childhood programs. As a result, measures of writing or its developmental precursors are not included here.

3. Correlations among measures were also calculated for Time 2 assessments, but results substantially replicate those reported here. Time 2 correlations are available upon request from either author.

References

Adams, M. (1990). *Beginning to read: Thinking and learning about print*. Cambridge, MA: Massachusetts Institute of Technology Press.

Anderson, R. C., Hiebert, E. H., Scott, J. A., & Wilkinson, I. A. G. (1985). *Becoming a nation of readers: The report of the commission on reading*. Washington, DC: National Institute of Education.

Assessment Committee Analysis of Reading Assessment Measures. (2002). Dynamic indicators of basic early literacy skills: Technical report. Retrieved from http://dibels.uoregon.edu/techreports/dibels_5th_ed.pdf

Bijou, S. W., Peterson, R. F., & Ault, M. H. (1968). A method to integrate descriptive and experimental field studies at the level of data and empirical concepts. *Journal of Applied Behavior Analysis*, *1*, 175–191.

Bliss, S., Skinner, C. H., & Adams, R. (2006) Enhancing an English language learning fifth-grade student's sight-word reading with a time-delay taped-words intervention. *School Psychology Review*, *35*(4), 663–670.

Burchinal, M., Vandergrift, B., Pianta, R. C., & Mashburn, A. (2010). Threshold analysis of association between child care quality and child outcomes for low-income children in pre-kindergarten programs. *Early Childhood Research Quarterly*, *25*, 166–176.

Burns, M. K., Haegele, K., & Petersen-Brown, S. (2014). Screening for early reading skills: Using data to guide resources and instruction. In R. J. Kettler, T. A. Glover, C. A. Albers, & K. A. Feeney-Kettler (Eds.), *Universal screening in educational settings: Evidence-based decision making for schools* (pp. 171–197). Washington, DC: American Psychological Association.

Cadigan, K. C., & Missall, K. N. (2007). Measuring expressive language growth in young children with autism spectrum disorders. *Topics in Early Childhood Special Education*, *27*(2), 110–118.

Carta, J., Greenwood, C., Walker, D., & Buzhardt, J. (Eds.). (2010). *Using IGDIs: Monitoring progress and improving intervention for infants and young children*. Baltimore, MD: Brookes.

Carta, J. J., Greenwood, C. R., Luze, G. J., Cline, G., & Kuntz, S. (2004). Developing a general outcome measure of growth in social skills for infants and toddlers. *Journal of Early Intervention*, *26*, 91–114.

Carta, J. J., Greenwood, C. R., Walker, D., Kaminski, R. A., Good, R., McConnell, S. R., & McEvoy, M. (2005). Individual Growth and Development Indicators (IGDIS): Assessment that guides intervention for young children. *Young Exceptional Children*, *4*, 15–27.

Chard, D. J., & Kame'enui, E. J. (2000). Struggling first-grade readers: The frequency and progress of their reading. *Journal of Special Education*, *34*(1), 28–38.

Christ, T. J., & Nelson, P. M. (2014). Developing and evaluating screening systems: Practical and psychometric considerations. In R. J. Kettler, T. A. Glover, C. A. Albers, & K. A. Feeney-Kettler (Eds.), *Universal screening in educational settings: Evidence-based decision making for schools* (pp. 79–110). Washington, DC: American Psychological Association.

Cronbach, L. J. (1990). *Essentials of psychological testing* (5th ed.). New York, NY: Harper & Row.

Cooper, B. R., & Lanza, S. T. (2014). Who benefits most from Head Start? Using latent class moderation to examine differential treatment effects. *Child Development*, *85*, 2317–2338.

DeCarlo, L. T. (1997). On the meaning and use of kurtosis. *Psychological Methods*, *2*, 292–307.

Deno, S. L. (1985). Curriculum-based measurement: The emerging alternative. *Exceptional Children*, *52*(3), 219-232.

Deno, S. L. (1986). Formative evaluation of individual student programs: A new role for school pschologists. *School Psychology Review*, *15*(3), 358-374.

Deno, S. L. (1997). Whether thou goest. . . Perspectives on progress monitoring. In J. W. Lloyd, E. J. Kame'enui, & D. Chard (Eds.), *Issues in educating students with disabilities* (pp. 77–99). Mahwah, NJ: Erlbaum.

Deno, S. L., Mirkin, B., & Chiang, P. K. (1982). Identifying valid measures of reading. *Exceptional Children*, *49*, 36–45.

Dickinson, D. K., & Neuman, S. B. (Eds.). (2006). *Handbook of early literacy research* (Vol. 2). New York, NY: Guilford.

Dolch, E. (1948). *Problems in reading*. Champaign, IL: Garrard Press.

Early Childhood Research Institute on Measuring Growth and Development. (1998). *Selection of general growth outcomes for children between birth and age eight* (Tech. Rep. No. 2). Minneapolis, MN: Center for Early Education and Development.

Early Childhood Research Institute on Measuring Growth and Development. (2004). *Psychometric characteristics of individual growth and development indicators: Picture naming, rhyming and alliteration* (Tech. Rep. No. 8). Minneapolis, MN: Center for Early Education and Development.

Espin, C., McMaster, K. L., Rose, S. A., & Wayman, M. M. (Eds.). (2012). *A measure of success: The influence of curriculum-based measurement on education*: Minneapolis: University of Minnesota Press.

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

Fuchs, L. S. (2003). Assessing intervention responsiveness: Conceptual and technical issues. *Learning Disabilities Research & Practice*, *18*(3), 172–186.

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

Fuchs, L. S., & Fuchs, D. (2007). The role of assessment in the three-tier approach to reading instruction. In D. Haager, J. Klingner, & S. Vaughn (Eds.), *Evidence-based reading practices for response to intervention* (pp. 29–44). Baltimore, MD: Brookes.

Goffreda, C. T., & Diperna, J. C. (2010). An empirical review of psychometric evidence for the Dynamic Indicators of Basic Early Literacy Skills. *School Psychology Review*, *39*, 463.

Good, R., & Jefferson, G. (1998). Contemporary perspectives on curriculum-based measurement validity. In M. R. Shinn (Ed.), *Advanced applications of curriculum-based measurement* (pp. 61–88). New York, NY: Guilford Press.

Good, R. H., & Kaminski, R. A. (1996). Assessment for instructional decisions: Toward a proactive/prevention model of decision-making for literacy skills. *School Psychology Quarterly*, *11*, 326–336.

Good, R. H., Gruba, J., & Kaminski, R. A. (2002). Best practices in using Dynamic Indicators of Basic Early Literacy Skills (DIBELS) in an outcomes-driven model. In A. Thomas & J. Grimes (Eds.), *Best practices in school psychology* (4th ed., Vol. 1, pp. 699–720). Washington, DC: National Association of School Psychologists.

Greenwood, C., Bradfield, T., Kaminski, R., Linas, M., Carta, J., & Nylander, D. (2011). The response to intervention approach in early childhood. *Focus on Exceptional Children*, *43*(9), 1–22.

Greenwood, C. R., Carta, J. J., & McConnell, S. R. (2011). Advances in measurement for universal screening and individual progress monitoring of young children. *Journal of Early Intervention*, *33*, 254–267.

Greenwood, C. R., Luze, G. J., & Carta, J. J. (2002). Best practices in assessment and intervention results with infants and toddlers. In A. Thomas & J. Grimes (Eds.), *Best practices in school psychology* (4th ed., Vol. 2, pp. 1219–1230). Washington, DC: National Association of School Psychologists.

Hojnoski, R. L., & Missall, K. N. (2006). Addressing school readiness: Expanding school psychology in early education. *School Psychology Review*, *35*, 602–613.

Hoover, W. A., & Gough, P. B. (1990). The simple view of reading. *Reading and Writing*, *2*(2), 127–160.

Hopkins, K. D., & Weeks, D. L. (1990). Tests for normality and measures of skewness and kurtosis: Their place in research reporting. *Educational and Psychological Measurement*, *50*, 717–729.

Juel, C. (2006). The impact of early school experiences on initial reading. In D. K. Dickinson & S. B. Neuman (Eds.), *Handbook of early literacy research* (Vol. 2, pp. 410–426). New York, NY: Guilford.

Kaminski, R. A., & Good, R. H. (1996). Toward a technology for assessing basic early literacy skills. *School Psychology Review*, *25*, 215–227.

Lindsey, K. A., Manis, F. R., & Bailey, C. E. (2003). Prediction of first-grade reading in Spanish-speaking English-language learners. *Journal of Educational Psychology*, *95*, 482–494.

Litwin, M. S. (2002). *How to assess and interpret survey psychometrics*. Thousand Oaks, CA: Sage.

Lonigan, C. J., Phillips, B. M., Clancy, J. L., Landry, S. H., Swank, P. R., Assel, M., . . . School Readiness Consortium. (2015). Impacts of a comprehensive school readiness curriculum for preschool children at risk for educational difficulties. *Child Development*, *86*, 1773–1793.

Luze, G. J., Linebarger, D. L., Greenwood, C. R., Carta, J. J., & Walker, D. (2001). Toward a technology of dynamic indicators of communicative expression for infants and toddlers. *School Psychology Review*, *30*, 393–416.

Marston, D. B., & Magnusson, D. (1988). Curriculum-based measurement: District level implementation. In J. L. Graden, J. E. Zins, & M. J. Curtis (Eds.) *Alternative educational delivery systems: Enhancing instructional options for all students* (pp. 137–172). Washington, DC: National Association of School Psychologists.

McConnell, S. R., & Missall, K. N. (2008). Best practices in monitoring progress for preschool children. In A. Thomas & J. Grimes (Eds.), *Best practices in school psychology* (5th ed., pp. 561–573). Washington, DC: National Association of School Psychologists.

McConnell, S., McEvoy, M., & Priest, J. (2002). "Growing" measures for monitoring progress in early childhood education: A research and development process for Individual Growth and Development Indicators. *Assessment for Effective Intervention*, *27*(4), 3–14.

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.

McConnell, S., & Wackerle-Hollman, A. (2006). *Advances in progress monitoring in early literacy*. Paper presented at the Conference on Advances in Progress Monitoring: Curriculum Based Measurement Research and Innovations, St. Paul, MN.

McConnell, S. R., Wackerle-Hollman, A. K., & Bradfield, T. A. (2014). Early childhood literacy screening. In R. Kettler, T. Glover, C. Albers, & K. A. Feeney-Kettler (Eds.), *Universal screening in educational settings: Identification, implications, and interpretation* (pp. 141–170). Washington, DC: American Psychological Association.

McMaster, K., & Espin, C. A. (2007). Technical features of curriculum-based measurement in writing: A literature review. *Journal of Special Education*, *41*, 68–84.

Missall, K. N. (2002). Reconceptualizing school adjustment: A search for intervening variables (Doctoral dissertation, University of Minnesota). *Dissertation Abstracts International: Section A*, *63*(5), 1712.

Missall, K. N., McConnell, S. R., & Cadigan, K. (2006). Early literacy development in preschool: Skill growth and relations between classroom variables for preschool children. *Journal of Early Intervention*, *29*, 1–21.

Missall, K. N., Reschly, A., Betts, J., McConnell, S. R., Heistad, D., Pickart, M., . . . Marston, D. (2007). Examination of the predictive validity of preschool early literacy skills. *School Psychology Review*, *36*(3), 433–452.

National Early Literacy Panel. (2009). *Developing early literacy: Report of the National Early Literacy Panel. A scientific synthesis of early literacy development and implications for intervention*. Jessup, MD: National Institute for Literacy.

National Reading Panel. (2000). *Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction*. Washington, DC: National Research Council.

Paris, S. G. (2005). Reinterpreting the development of reading skills. *Reading Research Quarterly*, *40*(2), 184–202.

Priest, J., Davis, S., McConnell, S., McEvoy, M., & Shinn, J. (1999). *Individual growth and development indicators of preschoolers' expressing meaning skills: Follow that trajectory!* Paper presented at the annual meeting of the Division for Early Childhood, Council for Exceptional Children, Washington, DC.

Priest, J. S., McConnell, S. R., Walker, D., Carta, J. J., Kaminski, R. A., McEvoy, M. A., . . . Shinn, M. R. (2002). General growth outcomes for children between birth and age eight: Where do you want young children to go today and tomorrow? *Journal of Early Intervention*, *24*, 163–168.

Priest, J. S., Silberglitt, B., Hall, S., & Estrem, T. L. (2000). *Progress on preschool IGDIs for early literacy*. Presentation at Heartland Area Education Association, Des Moines, IA.

Reifman, B., Pascarella, E. T., & Larson, A. (1981). Effects of word-bank instruction on sight word acquisition: An experimental note. *Journal of Educational Research*, *74*(3), 175–178.

Research Evaluation and Assessment. (2004). *Beginning kindergarten assessment*. Minneapolis, MN: Minneapolis Public Schools.

Rouse, H. L., & Fantuzzo, J. W. (2006). Validity of the Dynamic Indicators for Basic Early Literacy Skills as an indicator of early literacy for urban kindergarten children. *School Psychology Review*, *35*(3), 341–355.

Sénéchal, M., LeFevre, J.-A., Smith-Chant, B. L., & Colton, K. V. (2001). On refining theoretical models of emergent literacy: The role of empirical evidence. *Journal of School Psychology*, *39*(5), 439–460.

Snow, C. E., Burns, M. S., & Griffin, P. (Eds.). (1998). *Preventing reading difficulties in young children*. Washington, DC: National Academy Press.

Stanovich, K. E. (1988). Matthew effects in reading: Some consequences of individual differences in the acquisition of early literacy. *Reading Research Quarterly*, *21*, 360–407.

Tindal, G., Marston, D., & Deno, S. (1983). *The reliability of direct and repeated measurement* (Research Report No. 109). Minneapolis, MN: University of Minnesota, Institute for Research on Learning Disabilities.

Torgesen, J., & Bryant, B. (1994). *Test of Phonological Awareness*. Austin, TX: PRO-ED.

Walker, D., Greenwood, C., Hart, B., & Carta, J. (1994). Prediction of school outcome based on language production and socioeconomic factors. *Child Development*, *65*, 606–621.

Wallace, T. A., Espin, C. A., McMaster, K., Deno, S. L., & Foegen, A. (2007). CBM progress monitoring within a standards-based system: Introduction to the special series. *Journal of Special Education*, *41*, 66–67.

Wayman, M. M., Wallace, T. A., Wiley, H. I., Tichá, R., & Espin, C. A. (2007). Literature synthesis on curriculum-based measurement in reading. *Journal of Special Education*, *41*, 85–120.

Whitehurst, G. J., & Lonigan, C. J. (1998). Child development and emergent literacy. *Child Development*, *69*, 848–872.

Whitehurst, G. J., Zevenbergen, A. A., Crone, D. A., Schultz, M. D., Velting, O. N., & Fischel, J. E. (1999). Outcomes of an emergent literacy intervention from head start through second grade. *Journal of Educational Psychology*, *91*, 261–272.

Wilson, M. (2005). *Constructing measures: An item response modeling approach*. Mahwah, NJ: Erlbaum.

Woodcock, R., McGrew, K., & Mathers, K. (2002). *Woodcock-Johnson Test of Achievement–III*. River Falls, WI: Riverside.

Ysseldyke, J., Thurlow, M., Graden, J., Wesson, C., Algozzine, B., & Deno, S. (1983). Generalization from five years of research on assessment and decision making: The University of Minnesota Institute. *Exceptional Education Quarterly*, *4*(1), 75–93.

Authors

SCOTT MCCONNELL is professor of educational psychology and child psychology at the University of Minnesota, Room 351 Education Sciences Building (Mail Stop 4101A), 56 E River Parkway, Minneapolis, MN 55455. His research focuses on assessment and intervention of language and early literacy development, particularly for children with one or more developmental risk factors.

ALISHA WACKERLE-HOLLMAN is senior research associate in the Department of Educational Psychology at the University of Minnesota. Her research focuses on assessment of language and early literacy development of English- and Spanish-speaking students, as well development and evaluation of parent education programs for young families.