Predictive Validity of the *Get Ready to Read!* Screener

Concurrent and Long-Term Relations With Reading-Related Skills

Beth M. Phillips
Christopher J. Lonigan
Marcy A. Wyatt
Florida State University

This study examined concurrent and longitudinal relations for the *Get Ready to Read!* (GRTR) emergent literacy screener. This measure, within a battery of oral language, letter knowledge, decoding, and phonological awareness tests, was administered to 204 preschool children (mean age = 53.6, *SD* = 5.78; 55% male) from diverse socioeconomic backgrounds. Subgroups were reassessed at 6 months and 16 and 37 months later. Results indicate strong relations between the GRTR and the literacy and language assessments. Long-term follow-up indicated that the screener was significantly related to some reading-related measures, including decoding skills. These results support the utility of the GRTR as a brief, valid measure of children's emergent literacy skills. The GRTR holds promise as a tool useful for educators, parents, and others in regular contact with preschool children to help determine those who may be at risk for later reading difficulties and could benefit from intervention and focused instruction in emergent literacy.

Keywords: preschool age; identification; assessment; early literacy; reading

The past several decades have been witness to a notable increase in knowledge about the causes, correlates, and predictors of children's reading success and failure. An associated phenomenon is greater interest in early identification and intervention for children determined to be at risk for reading problems. In large measure, this stems from research suggesting that children's reading performance is highly stable from early in elementary school (e.g., Juel, 1988) and from research indicating more success with prevention and earlier intervention than with remediation for older students (Berninger et al., 2002; Coyne, Kame'enui, Simmons, & Harn, 2004; Torgesen, 2000). Of course, to successfully prevent reading difficulties for at-risk students, they first have to be identified accurately.

As the focus of federal and state educational policies and early learning standards has shifted toward the inclusion of early literacy skill development, the need has grown to identify constructs and reliable measures of these constructs that best predict who will and will not

develop significant reading difficulties or well-below-average reading ability. Several large-scale longitudinal studies have demonstrated that assessments of key skills, such as phonological awareness (PA), letter knowledge, rapid naming (RN), and oral language, conducted just before or during the onset of reading instruction (i.e., in kindergarten and first grade) are significant predictors of reading ability 1 or more years later (e.g., Catts, Fey, Zhang, & Tomblin, 1999; de Jong & van der Leij, 2003; Schatschneider, Fletcher, Francis, Carlson, & Foorman, 2004; Wagner, Torgesen, & Rashotte, 1994).

Whereas most of the predictive studies have first assessed children in kindergarten or just before, a few studies have investigated the predictive power of batteries administered to preschool children. For instance, studies by Badian (e.g., 1988, 1994, 1998) have shown that multidimensional assessments have adequate predictive power up to 9 years later (e.g., Badian, 1998, reported an 87% hit rate). Badian (1994) achieved good results with a battery that included demographic and

family history information as well as child performance on PA, RN, and visual matching tasks. Fowler and Cross (1986) used a combination of family characteristics and child data to predict reading and math performance for preschool children after they completed 2 to 3 years of formal schooling and found that a score of 7 or higher on their 21-point risk index had a 98% positive predictive power for successful grade completion. Studies of preschool children that have relied only on child assessments also have predicted successfully later reading abilities (e.g., Blatchford, Burke, Farquhar, & Plewis, 1987; Bowey, 1995; Chaney, 1998; Stevenson & Newman, 1986; Storch & Whitehurst, 2002).

For both kindergarten and preschool children, several published norm- or criterion-referenced measures have been created to serve as diagnostic assessment instruments. Examples include the *Test of Phonological Awareness–2* (Torgesen & Bryant, 2004), the *Test of Preschool Early Literacy* (TOPEL; Lonigan, Wagner, Torgesen, & Rashotte, 2007), the *Test of Early Reading Ability–3* (Reid, Hresko, & Hammill, 2001), the *Texas Primary Reading Inventory* (TPRI; Texas Education Agency, 1999), and the *Phonological Awareness Literacy Screening* in both kindergarten and preschool versions (PALS-K and PALS-PreK; Invernizzi, Juel, Swank, & Meier, 2007; Invernizzi, Sullivan, Meier, & Swank, 2004). These measures have demonstrated adequate psychometric properties and concurrent validity; however, few of them have established longitudinal predictive capacity in terms of either correlations with later decoding or comprehension measures or power to identify children at risk for actual reading disabilities (Havey, Story, & Buker, 2002).

An alternative possibility to criterion- or norm-referenced tests is kindergarten-entry "readiness tests" that typically include broad measures of cognitive and academic skills. For example, Duncan and Rafter (2005) found the *Phelps Kindergarten Readiness Scale* to be a moderate predictor of Woodcock-Johnson Achievement scores across an 8-month interval in kindergarten. Some of these tests were designed to determine whether children had adequate skill levels to begin school, a practice that, although still widespread, has not shown demonstrable validity (May & Kundert, 1997; Morrison, Griffith, & Alberts, 1997; Shepard, 1997; Stipek, 2002). Several reviews report widely varying success for these kindergarten entry measures in predicting later academic outcomes, including reading performance (e.g., La Paro & Pianta, 2000; Pianta & McCoy, 1997; Shepard, 1997; Tramontana, Hooper, & Selzer, 1988). La Paro and Pianta's (2000) meta-analysis of 32 studies predicting

academic outcomes in kindergarten or first grade from academic measures in preschool showed an average predictive correlation of .43, with a range from .08 to .72. Data from these studies and from the recently completed National Early Literacy Panel's predictive study (Lonigan, Schatschneider, & Westberg, in press) suggest that the measures that are more successful predictors of decoding skill typically include key emergent literacy skills, such as letter knowledge and PA, rather than general cognitive or developmental abilities (e.g., Bramlett, Rowell, & Mandenberg, 2000; Chew & Morris, 1989; Flynn & Rahbar, 1998; Lonigan, Schatschneider, & Westberg, in press).

Although findings from predictive studies are largely favorable and suggest that it is possible to identify children at risk for reading problems, most of the research batteries or clinical assessments are lengthy and time consuming and often require highly trained assessors or specialized equipment. This is also an issue for many, although not all, of the kindergarten-entry diagnostic assessments mentioned previously. These factors make such assessments prohibitively expensive and inefficient for school systems that want to screen large numbers of students each year. That is, center-based programs prefer to use informal teacher judgments or brief, universal survey measures initially that can be used to help identify those children who may benefit from receiving a longer, diagnostic assessment, as determined formally by a cutoff score indicating categorical risk status. Moreover, there is a significant issue of how to identify preschool children who may not be attending center-based educational or child care programs where a trained teacher could administer a measure. To be truly universally applicable, a screening tool would need to be usable by parents, pediatricians, and other professionals who come into regular contact with 3- and 4-year-old children.

Development of *Get Ready to Read!*

To help fill the need for a reliable, research-based screening measure for use with 3- to- 5-year-old children, Whitehurst and Lonigan developed the *Get Ready to Read!* (GRTR) screening tool in conjunction with the National Center for Learning Disabilities (NCLD; Whitehurst, 2001). The goal for development was to create a brief, user-friendly measure that had strong concurrent relations to more comprehensive measures of children's print knowledge, letter knowledge, and PA skills that themselves had established validity in predicting reading skill.

Items similar to those from measures used in prior research were used to create an initial pool of 100 candidate

items, all in a multiple-choice format using easily identified pictures of both target and alternate responses. All items required only a pointing response from the child, an asset when developing a measure designed to be administered by lay individuals. Elimination of redundancy and items that were deemed overly challenging led to the creation of a 60-item alpha pool. The 60 items assessed print concepts, letter-name and letter-sound knowledge, rhyming, initial sound matching, compound word blending, and emergent writing (Whitehurst, 2001). These items were then administered to cohorts of 3-, 4-, and 5-year-old children in preschool classrooms ($N = 342$ total). The children were representative of a wide array of socioeconomic backgrounds but were somewhat overrepresentative of lower-income students attending Head Start (i.e., $N = 223$), as these children are known to be at risk for later literacy problems. Most of the children (e.g., 80%) were 4-year-olds.

Analyses of internal consistency, item-total correlations, correlations with the *Developing Skills Checklist* (DSC; CTB/McGraw-Hill, 1990), and item difficulty were used to reduce the alpha pool to the final 20-item screener. This measure demonstrated a coefficient alpha of .78 and a split-half reliability of .80. The total score correlation with the DSC was .69. Children's scores showed a relatively normal distribution, with 68% of the children scoring between 5 and 13 correct, and scores from 1 to 20 were represented ($M = 9.14$, $SD = 4.31$). GRTR scores were somewhat correlated with age, although visual and statistical analyses suggested that this correlation was driven mostly by the 3- and 5-year-old children in the sample (i.e., the correlation was reduced and nonsignificant within the exclusively 4-year-old sample). Furthermore, reliability and validity coefficients for subsamples of middle- and lower-income children were highly comparable (i.e., split-half reliability of .78 and .80 for middle- and lower-income samples, respectively), with only mean scores differing significantly. Principal components analysis indicated that one factor best described the screening measure in its current 20-item form (Whitehurst, 2001).

Since its development, NCLD has conducted several large-scale demonstration projects involving dissemination of the GRTR and evaluation of growth in scores from the beginning to the end of the preschool year. Data from one such project, a study that included approximately 1,200 4-year-old children attending preschools in Arizona, Georgia, and Maryland, indicated that the average raw score in the fall was 13.14 whereas the average score at the end of the year was 16.14, a gain of 15% on average (NCLD, 2003). In this study, about two thirds of the children achieved scores equal to or higher than 16

out of 20 on the end-of-year administration. According to Whitehurst (2001), scores in this range are indicative of strong skills.

To date, the only published studies using the GRTR tool were reported by Molfese and colleagues (e.g., Molfese et al., 2006; Molfese, Molfese, Modglin, Walker, & Neamon, 2004). The 2004 study was a concurrent validity assessment of the relations between the GRTR and measures of general cognitive ability, expressive and receptive vocabulary, rhyming, blending, and environmental print. The sample included 152 3- and 4-year-old children attending subsidized preschool programs for children from economically disadvantaged families. Results indicated that the 4-year-old children scored higher, on average, than did the 3-year-old children on all measures except expressive vocabulary; GRTR mean scores were 7.9 and 11.51 for 3- and 4-year-old children, respectively. The GRTR had statistically significant and moderate to large correlations with all measures except the blending task for 4-year-old children, and with all measures except environmental print for 3-year-old children. In the later study, Molfese et al. (2006) found that the GRTR administered in the spring significantly correlated with fall to spring gains in letter knowledge as well as with both fall and spring letter knowledge total scores. GRTR scores also differentiated children in the high- and low-letter knowledge growth subgroups.

Whereas existing data on the GRTR are promising and suggest that the GRTR measures constructs relevant to early literacy skills in a reliable and valid way, the ultimate test of a screening tool's utility is how well it predicts later success on reading-related measures. Therefore, the goal of this study was to evaluate the short- and long-term predictive relations of the 20-item GRTR measure both to more detailed measures of emergent literacy skills and to measures of actual reading ability. These analyses are conceptualized as an intermediary goal in the process of determining whether the GRTR is appropriate for use as a formal screening measure, with a cutoff score establishing risk for poor performance on a larger diagnostic measure. Specifically, the research questions were as follows: (a) What are the concurrent validity relations between the GRTR and longer single-focus measures of language, print, decoding, and phonological ability? (b) What are the short-term predictive validity relations between the GRTR and these established measures? and (c) What is the longer term predictive validity of the GRTR with respect to measures of language, foundational literacy skills such as letter knowledge and phonological awareness, and actual decoding and reading comprehension skill?

Method

Participants

Original sample. All participants were randomly selected for participation in this project from three longitudinal assessment projects ongoing in fall 2000. Participants were selected for inclusion if they were 3- to 5-years-old at the time and were attending a Head Start center (41.2%), public prekindergarten (33.8%), or a private preschool center (25%). Fewer than 11% of these children were exposed to a research-based curriculum or intervention specifically targeting early literacy skill development, although their various classroom curricula did sometimes include literacy-focused activities. The GRTR was completed by 204 children who represented a range of socioeconomic backgrounds. Children ranged in age from 37 to 63 months ($M = 52.15$, $SD = 5.57$) and included 113 boys (55.4%). Whereas some of the children were designated by their schools as having mild to moderate language delays or other mild disabilities (this was a condition of preferential admission for the public preschool programs), the sample did not include any children who were known to have moderate to severe developmental disabilities (as reported by teachers or as observed by assessors). Consistent with the desire to have a sample overrepresentative of children who may be at risk for reading difficulties because of poverty and associated factors and with the fact that in the local area, the majority of lower-socioeconomic status (SES) families are minorities, the sample included 65.7% African American children, 30.9% Caucasian children, and 3.4% children of other ethnic backgrounds. Although no exclusionary rules were in place with regard to home language, consistent with these ethnic backgrounds, it is estimated that less than 1% of the children were not native English speakers as observed by the field assessors. Parent report of home language was not collected.

Follow-up samples. Between 3 and 7 months ($M = 5.6$ months, $SD = 0.86$) following the initial assessment, 159 of the original participants were available for assessment. Four participants were removed from analyses because of excessive missing data, leaving a final sample of 155 children (76% of original sample). At baseline, these children were 38 to 62 months old ($M = 52.17$, $SD = 5.50$) and included 85 boys (54.8%). Their age range at follow-up was 43 to 69 months. As with the original sample, the majority of these children were African American (74.8%), 22.6% were Caucasian, and 2.6% represented other ethnic backgrounds. This sample did not differ significantly from the larger baseline sample in age or gender but did differ in ethnicity such that there was a greater representation of African American children within the short-term follow-up sample, $\chi^2(2, N = 204) = 23.99$, $p < .001$. This subsample did not differ significantly

from the larger group on average screener score or any baseline measure except receptive vocabulary (i.e., full sample: $M = 81.11$, $SD = 17.44$; short-term follow-up sample: $M = 79.48$, $SD = 17.29$), $F(1, 203) = 5.76$, $p < .05$.

A subset of the original sample was available for long-term follow-up assessment 16 to 37 months later. The wide range of follow-up intervals resulted from these children being part of three separate longitudinal study cohorts that had different follow-up timelines and between one and three waves of follow-up. Moreover, these varying intervals and the smaller sample size are indicative of the transitory living circumstances of many of the children from lower-SES families who participated. It was often very difficult to acquire and maintain contact with their families to arrange for follow-up assessment once they left their preschool centers. Some children not included at a first follow-up wave because they were not locatable may have been found later and included for a second follow-up wave. Data for this project are not demarcated by wave, although length of interval is noted. A small subset of children participated in assessments during multiple waves of follow-up; to provide larger sample sizes on some of the most relevant measures, we included only the first wave of follow-up data for these participants.

Overall, this long-term follow-up sample included 114 children (56% of the original sample). At the time of screening, these children were 39 to 63 months old ($M = 53.86$, $SD = 5.55$) and included 68 boys (59.6%). At the time of follow-up, the children ranged in age from 58 to 98 months ($M = 80.04$, $SD = 8.66$), with intervals between assessments ranging in length between 16 and 37 months ($M = 26.18$, $SD = 6.02$). Comparable to the full baseline sample, this subset included 64.9% African American children, 30.7% Caucasian children, and 4.4% children of other ethnicities. This sample did not differ significantly from the larger baseline sample in age, gender, or ethnicity. When compared on baseline data, the long-term follow-up sample was equivalent to the full sample on all measures (i.e., all *p*s > .05). Moreover, the sample retained the original proportionality of including approximately two children from a lower-income background for every child from a middle-income background.

Measures

Baseline and Younger Child Follow-Up Measures

Get Ready to Read! screener. The GRTR, as was described above, is a 20-item multiple-choice measure

that includes items measuring letter-name and letter-sound knowledge (5 items), PA (7 items), print concepts (5 items), and emergent writing knowledge (3 items). Each item, including a sample item, includes an orally presented question (e.g., point to the letter that makes the /b/ sound) and a picture page with four choices, including the target response and three foils (i.e., B, L, K, S). Children were instructed to respond by pointing to their answer choice. Development study findings indicated that the measure represents a single factor; therefore, only total scores were computed for analyses. At baseline, participating children were administered the GRTR in one-to-one sessions in their preschool centers. Administration time was approximately 5 minutes per child. Cronbach's alpha for the full initial sample ($N = 204$) was .79, and it was .78 for the subgroup that participated in the short-term follow-up and .71 for the subsample that participated in the long-term follow-ups.

Oral language measures. The *Peabody Picture Vocabulary Test–Revised (*PPVT-R; Dunn & Dunn, 1981) and the *Expressive One-Word Picture Vocabulary Test–Revised* (EOWPVT-R; Gardner, 1990) were administered at baseline and at long-term follow-up. Manuals for both measures report high reliability and validity for this age group.

Decoding measures. All children completed the *Word Identification* (WID) and *Word Attack* (WA) subtests of the *Woodcock Reading Mastery Test–Revised* (WRMT-R; Woodcock, 1987) at baseline and one or more follow-ups. For WID, children were asked to read aloud a series of age-appropriate real words; likewise, for WA, children were asked to read aloud a series of pronounceable nonwords. All children also completed a task in which they had to read 15 high-frequency words (Frequent Words) printed on large index cards. All items were scored 1 (*correct*) or 0 (*incorrect*); no credit was given if the child only named the letters or spelled the word. This measure was administered at all assessment points. Raw scores were used for these decoding measures at all time points because Frequent Words is not a standardized measure and because we have administered the WRMT measures to children younger than the standardization sample as a means of having continuity of measurement for decoding skill.

Print knowledge. Children were shown individual cards depicting 25 uppercase letters (due to a clerical error, the letter W was not included in the assessment) and asked to say the letter name. Letters were shown to the children in a standardized, nonalphabetical order. Administration stopped if the child failed to correctly name 5 consecutive letters. Immediately following this task, children were shown 8 of the letters (i.e., M, B, D, A, C, O, P, S) and asked to say what sound the letter made. Children were shown these letters regardless of whether they had been able to state the letter's name. If the child responded with the letter name, they were told, "That's right, but what sound does it make?" and were not given credit for naming the letter or saying a word that began with the letter. All letter-name and letter-sound items were scored 1 (*correct*) or 0 (*incorrect*). At baseline and long-term follow-up, children also completed an adapted version of Clay's (1979) *Concepts About Print Test* on which they were asked to identify common features of a book (e.g., front cover, print is read left to right) and name several punctuation marks. Only children younger than 84 months at the time of follow-up assessment completed these measures, as most children at or near that age score at or close to ceiling on these measures.

Phonological awareness tasks. All children younger than 84 months at the time of their baseline assessment or any follow-up completed a group of eight PA tasks. These tasks represented an early research version of the PA subtest on the TOPEL (Lonigan et al., 2007). Children completed three blending tasks that assessed the child's PA by asking him or her to blend words, syllables, and phonemes to create real words (e.g., "What word do you get when you say cow—boy together?" "What word do you get when you say /m/—/oo/—/n/ together?"). Children also completed three elision tasks in which they had to remove phonemes, syllables, or half of a compound word to state what word remained (e.g., "Say *sunshine*. Now say *sunshine* without saying *sun*." "Say *candy*. Now say *candy* without saying *dee*."). Children completed two tasks involving rhyming words, one in which they had to identify which of two choices rhymed with a presented word, and one in which they had to identify which of three presented words did not rhyme with a presented word. Four of these eight tasks involved the use of pictures for some or all words to reduce the memory load of the tasks. Within each skill area (i.e., rhyming, blending, elision), the relevant tasks were summed to create composite scores. Internal consistency coefficients for these measures in prior research ranged from adequate to excellent (e.g., average alpha coefficients for each composite were .59 for rhyming, .82 for blending, and .75 for elision; Lonigan, Anthony, et al., in press). Prior investigations using these and similar measures indicate significant correlations with measures of RN, phonological memory (PM), and letter knowledge, as well as with decoding tasks (e.g., Lonigan, Anthony, et al., in press; Lonigan, Burgess, & Anthony, 2000).

Rapid naming tasks. At baseline and short-term follow-up assessments only, children were administered three RN task versions in which they had to orally label four familiar objects arrayed in random order in six rows of four on a single page. Across the three versions, the four pictures rhymed (i.e., cat, bat, hat, rat), did not

rhyme (dog, ball, man, tree), or were two sizes of circles or squares that the children had to identify as big or little. Children practiced naming each item before timing began, and they completed two trials for each array. Scoring was the seconds required to name all objects, and the score used here represents the average naming time across the three arrays. In a recent large study of more than 400 2- to 5-year-old children, these RN tasks had strong test-retest reliabilities ranging from .82 to .85 (Lonigan, Anthony, et al., in press).

Whereas the measures of PA and lexical access (i.e., RN) abilities used in this study have face validity for the construct they are intended to assess, use of these or similar measures with other preschool samples provides evidence for their predictive convergent and discriminant validity. In a sample of 100 preschool and kindergarten children (*M* age = 68.0 months, *SD* = 11.12) who completed the measures of PA and lexical access abilities used in this study and the *Comprehensive Test of Phonological Processing* (CTOPP; Wagner, Torgesen, & Rashotte, 1999) 18 to 24 months later, partial correlations (controlling for age at preschool testing) between composite PA ($r = .31$) and lexical access ($r = .31$) measures with CTOPP PA and lexical access subtest standard scores, respectively, were statistically significant for within-construct longitudinal correlations and not significant for between-construct longitudinal correlations (average $r = .08$), and the longitudinal correlations within construct were significantly higher than the longitudinal correlations across constructs.

Older Child Follow-Up Measures

Children who were at least 60 months of age or older (to minimize floor effects with younger children) at the time of their long-term follow-up were administered some or all of the measures described below instead of or in addition to the previously described blending, elision, and RN measures. Decision rules based on ages (i.e., 60 months or older, younger than 84 months) were used within these projects to maximize the likelihood that children would provide a score for each construct of interest (e.g., PA) that was not at floor or ceiling levels. Thus, children between these two decision marks may have received both the measures designed for younger children (e.g., the eight PA tests) and the measures designed for older children (e.g., the CTOPP subtests). Including children who received either one or both of the measurement sets for each construct in the analyses maximized the sample sizes available at follow-up.

Comprehensive Test of Phonological Processing. The CTOPP (Wagner et al., 1999) assesses skills in all three domains of phonological processing—PA, RN, and PM—and composite standard scores are available for each domain. Between six and eight subtests of the CTOPP were administered to older children at their long-term follow-up, depending on their age and the project from which they were drawn. Depending on the subtests administered, the test yields up to three composite standard scores for PA, RN, and PM. Internal consistency reliability for the three composites ranges from .83 to .96 for children age 5 years and older, and 2-week test-retest coefficients range from .79 to .86. Validity for the CTOPP is supported by significant concurrent and predictive relations with the WRMT-R, the *Wide Range Achievement Test–Third Edition*, and the *Gray Oral Reading Test–Third Edition*.

Test of Word Reading Efficiency (TOWRE). The TOWRE (Torgesen, Wagner, & Rashotte, 1999) includes two subtests, each of which require a child to accurately read aloud as many words (Sight Word Efficiency) or pronounceable nonwords (Phonetic Decoding Efficiency) as he or she can within 45 seconds. The TOWRE is normed for children starting at age 6 years and yields several types of standard scores, normal curve equivalents, and percentile ranks, and an overall standard score combines the results of both subtests. The concurrent correlation of Sight Word Efficiency with the WRMT-R WID was .89, whereas the correlation between Phonetic Decoding Efficiency and the WRMT-R WA was .85 (Torgesen, Wagner, Herron, & Rashotte, 1998). For this study, only the combined total word reading standard score (TOWRE-SS) was used. The TOWRE was administered only at long-term follow-up to children older than 60 or 84 months, depending on the particular project's protocol (see Table 1).

Passage Comprehension (PC). A subset of children completed PC from the WMRT-R at long-term follow-up. This measure has adequate reliability and validity within this age group (Pae et al., 2005; Woodcock, 1987). Because some of the children who received this measure were younger than the standardization sample, normative standard scores were not used.

Gray Oral Reading Test–4th Edition (GORT-4). A subset of the participants received the GORT-4 at long-term follow-up assessment. The GORT-4 (Wiederholt & Bryant, 2001) yields scores of reading rate, accuracy, fluency, and comprehension appropriate for children age 6 years to 18 years. An overall reading ability standardized Oral Reading Quotient (ORQ) combines all subcomponent standard scores. All coefficient alphas for the subscores and ORQ across all ages and student subgroups exceed .90, and the average alternate form and test-retest reliabilities for the four subscores all exceed .85 (Wiederholt & Bryant, 2001). Substantial concurrent and predictive relations with other established measures of reading and related abilities

Table 1
Description of Assessment Protocol Across Time Points

Measure	Baseline (N = 204)	Short-Term Follow-Up (N = 155)	Long-Term Follow-Up (N = 114)
GRTR	x	—	—
PPVT-R	x	—	x
EOWPVT-R	x	—	x
3 blending measures	x	x	x; if < 84 months
3 elision measures	x	x	x; if < 84 months
2 rhyming measures	x	x	x; if < 84 months
Letter Name	x	x	x; if < 84 months
Letter Sound	x	x	x; if < 84 months
RN	x	x	—
Concepts About Print	x	—	x; if < 84 months
Word Identification	x	x	x
Word Attack	x	—	x
Frequent Words	x	x	x
CTOPP-PA	—	—	x; if ≥ 60 months
CTOPP-RN	—	—	x; if ≥ 60 months
CTOPP-PM	—	—	x; if ≥ 60 months
TOWRE	—	—	x; if ≥ 60 months
Passage Comp.	—	—	x; if ≥ 60 months
GORT-4	—	—	x; if ≥ 60 months

Note: GRTR = *Get Ready to Read!* screener; PPVT-R = *Peabody Picture Vocabulary Test–Revised*; EOWPVT-R = *Expressive One-Word Picture Vocabulary Test–Revised*; CTOPP = *Comprehensive Test of Phonological Processing*; PA = Phonological Awareness; RN = Rapid Naming; PM = Phonological Memory; TOWRE = *Test of Word Reading Efficiency*; Passage Comp. = Passage Comprehension subtest of the *Woodcock Reading Mastery Test*; GORT-4 = *Gray Oral Reading Test–4th Edition*.

demonstrated the validity of the GORT-4. Only the ORQ scores were used in these analyses.

Procedures

Parental consent was received initially for the baseline and short-term follow-up and then reacquired for the one or two long-term follow-up assessments. Trained research assistants with experience assessing literacy and language in these age groups conducted the assessments at the children's preschool centers. Training involved extensive opportunities for observing model administrations, administering the assessments to the master trainers and receiving feedback, and conducting practice child assessments. Long-term follow-ups were arranged individually with parents and were conducted in the children's homes or elementary schools, or in the laboratory. The GRTR was administered in the late fall of the baseline year, approximately 2 months after the other baseline data were collected for the ongoing larger projects. All assessment measures were administered in two to five 15- to 30-minute assessment sessions, depending on the size of the battery being given at that assessment wave. Whereas corrective feedback was not given, examiners provided nondifferential feedback focused on encouraging effort, attention, and motivation during and between all tasks.

Results

Within the full baseline sample, 12 children were missing data on FW and WA due to faulty administration procedures. For the PPVT-R, CTOPP, EOWPVT, TOWRE, and GORT-R, norm-referenced standard scores were used in all analyses. For all other measures, age-standardized scores were used in the analyses unless otherwise indicated. All age standardization occurred within the time interval by regressing raw scores on child age at the time of assessment and retaining the standardized residual. Use of these age-standardized scores represents a conservative way of analyzing the data, as it takes into account the range of ages for children when they received both the screener and their follow-up assessments. Age and GRTR were moderately correlated at .47 ($p < .001$). Following Tabatchnick and Fidell (2001), significant outliers (i.e., scores more than 3 standard deviations from the mean) were recoded to ±3.00 standard deviations. Descriptive statistics for the baseline assessment are included in Table 2 for both the baseline sample of 204 children and the 155 children who provided short-term follow-up data. For both of these samples, the mean GRTR scores of approximately 10 were somewhat lower than the fall scores of the large 2003 NCLD study sample. This result was likely a function of both the inclusion of 3-year-olds in the current sample and the high representation of children from lower-SES backgrounds.

Descriptive statistics for the short-term follow-up are included in Table 2. Repeated measures analyses of variance (ANOVAs) indicated that over the approximately half-year interval, children gained substantially in their PA abilities, with significant growth on blending, elision, and rhyming measures (e.g., all $ps < .001$). Likewise, children gained significantly in their letter-name and letter-sound knowledge, such that the average number of letter names known grew from approximately 7 to approximately 12, and the average number of letter sounds known increased from less than one to roughly two (both $ps < .001$).

The initial question addressed by these analyses was whether the GRTR measure was concurrently related to other measures of emergent literacy at the baseline assessment (see Table 3). Results for the conservative

Table 2
Means and Standard Deviations (in parentheses) for Baseline and Short-Term Follow-Up Data

Measure (maximum possible)	Full Sample (N = 204) Baseline	Short-Term Follow-Up Sample (N = 155) Baseline	Short-Term Follow-Up Sample (N = 155) Follow-Up
GRTR (20)	10.12 (4.45)	9.85 (4.39)	—
PPVT-R-SS	81.11 (17.44)	79.48 (17.29)	—
EOWPVT-SS	86.35 (13.26)	85.85 (13.71)	—
Blending composite (31)	12.56 (5.97)	12.64 (5.93)	18.03 (7.13)
Elision composite (31)	6.88 (4.90)	6.56 (4.63)	13.27 (6.79)
Rhyme composite (22)	7.81 (3.79)	7.53 (3.67)	10.88 (5.24)
Letter Name (25)	7.36 (8.99)	6.81 (8.83)	12.02 (10.25)
Letter Sound (8)	0.82 (1.40)	0.75 (1.39)	2.03 (2.72)
Rapid Naming (sec)	56.29 (21.43)	56.75 (22.23)	43.39 (11.78)
Concepts About Print (13)	4.15 (2.53)	3.97 (2.55)	—
Word Identification (71)	0.10 (0.74)	0.12 (0.84)	0.23 (0.84)
Word Attack (20)	0.01 (0.14)	0.01 (0.16)	—
Frequent Words (15)	0.06 (0.29)	0.04 (0.23)	0.24 (0.89)

Note: *N* = 192 for baseline Frequent Words and Word Attack measures. GRTR = *Get Ready to Read!* screener; PPVT-R-SS = *Peabody Picture Vocabulary Test–Revised–Standard Score*; EOWPVT-SS = *Expressive One-Word Picture Vocabulary Test–Standard Score*.

age-standardized analyses indicate that the GRTR was significantly correlated with all measures administered at baseline except WA, for which there was a substantial floor effect (i.e., significant correlations ranged from .18 for Frequent Words to .61 for Letter Name, all *p*s < .05). These findings held virtually identically regardless of whether the full baseline or the short-term follow-up sample was used. Furthermore, no meaningful differences in the overall pattern of correlations were found when analyses were restricted to White, African American, and male or female participants.

The second question addressed was whether the 20-item GRTR screener was related to emergent literacy data collected at the short-term follow-up. The short-term longitudinal correlations for the GRTR with other measures are shown in Table 3. These data indicate that across this 3- to 7-month interval, the GRTR was significantly correlated with all assessments administered. To confirm that these findings were not a function of the range of interassessment intervals or of the range of children's age at baseline, these data also were analyzed excluding the small number of participants (e.g., 4.5%) with follow-up intervals less than 5 months and excluding participants younger than 48 months at baseline. No changes were found to the predictive relations between the GRTR and the criterion variables. Also noteworthy is the fact that despite including only a small number of items measuring each aspect of PA, the 20-item GRTR achieved a cross-time correlation comparable to that of the longer, more detailed criterion measure with itself (e.g., the cross-time

Table 3
Concurrent and Longitudinal Correlations of *Get Ready to Read!* Screener With Age-Standardized Baseline and Short-Term Follow-Up Data for Full and Short-Term Follow-Up Samples

Measure	Full Sample Baseline	Short-Term Sample Baseline	Short-Term Sample Follow-Up
PPVT-R-SS	.46***	.45***	—
EOWPVT-SS	.51***	.52***	.37***
Blending composite	.46***	.45***	.32***
Elision composite	.36***	.29***	.40***
Rhyming composite	.36***	.30***	.25***
Letter Name	.61***	.60***	.36***
Letter Sound	.34***	.29***	.34***
Rapid Naming	−.36***	−.39***	−.25**
Concepts About Print	.31***	.36***	—
Word Identification	.21**	.24**	.32***
Word Attack	.14	.16	—
Frequent Words	.18*	.12	.25***

Note: *N* = 204 for full sample data; *N* = 155 for short-term sample data. PPVT-R-SS = *Peabody Picture Vocabulary Test–Revised–Standard Score*; EOWPVT-SS = *Expressive One-Word Picture Vocabulary Test–Standard Score*.
*$p < .05$. **$p < .01$. ***$p < .001$.

autocorrelation for the Elision Composite was .39 [*p* < .001] as compared with .40 [*p* < .001] for the GRTR with the measure at short-term follow-up).

The final research question addressed the longer-term predictive validity between the GRTR and measures of

Table 4
Means and Standard Deviations (in parentheses)
for Longitudinal Follow-Up Cohort

Measure (maximum possible)	Assessment Point	
	Baseline	Follow-Up
GRTR-School (20)	9.74 (3.86)	—
PPVT-R-SS	81.43 (17.36)	91.70 (17.46)
EOWPVT-SS	87.33 (13.91)	100.21 (22.18)
Blending composite (31)	12.88 (5.69)	25.63 (6.47)
Elision composite (31)	6.83 (4.85)	21.33 (6.98)
Rhyme composite (22)	7.76 (3.53)	18.25 (3.59)
Concepts About Print (13)	4.34 (2.61)	9.46 (3.55)
CTOPP-PA-SS	—	99.73 (13.93)
CTOPP-RN-SS	—	97.63 (12.88)
CTOPP-PM-SS	—	98.11 (13.68)
Letter Name (25)	6.63 (8.34)	23.16 (5.25)
Letter Sound (8)	0.68 (1.12)	6.62 (2.43)
Word Identification (71)	0.13 (0.94)	25.85 (21.58)
Word Attack (20)	0.00 (0.00)	9.93 (10.50)
Frequent Words (15)	0.71 (0.32)	5.94 (5.04)
TOWRE-SS	—	100.63 (14.20)
Passage Comprehension (68)	—	16.55 (10.47)
GORT-4 ORQ	—	82.17 (18.72)

Note: Sample sizes at follow-up ranged from 42 to 111. GRTR = *Get Ready to Read!* screener; PPVT-R-SS = *Peabody Picture Vocabulary Test–Revised–Standard Score*; EOWPVT-SS = *Expressive One-Word Picture Vocabulary Test–Standard Score*; CTOPP = *Comprehensive Test of Phonological Processing*; PA = Phonological Awareness; RN = Rapid Naming; PM = Phonological Memory; SS = Standard Score; TOWRE-SS = *Test of Word Reading Efficiency–Standard Score*; GORT-4 = *Gray Oral Reading Test–4th Edition*; ORQ = Oral Reading Quotient.

reading (e.g., decoding and comprehension) and reading-related skills (e.g., phonological processing). To maximize the sample available for these analyses, all children for whom scores on a measure were available were included in that particular analysis. Descriptive data for children included in the longitudinal follow-up cohort are shown in Table 4. Both baseline and follow-up data are included for all measures administered at the relevant time points. Whereas the full group of 114 participants received all baseline measures, varying subsets of children received some or all of the measures at follow-up, determined by the child's age at the assessment interval and the longitudinal project in which he or she was a participant. For example, a child who participated in follow-up at a 20-month interval but was younger than 84 months would not have received the TOWRE. As noted above, these age criteria were used to minimize the likelihood that a child would score at floor (or ceiling) levels on measures of varying difficulty and to likewise minimize children's sense of frustration during administration.

Follow-up measures for which at least some participants provided data are included in analyses. For all but two measures (i.e., PC and GORT-ORQ, for which there were 42–44 participants), there were at least 64 participants in the analyses with the GRTR.

Separate regression and partial correlation analyses were conducted for each follow-up measure. Within all of these regression analyses, the child's age at the time of screening and the length of the interval (in months) between baseline and follow-up assessment were included as covariates. Likewise, the partial correlations were conducted controlling for the child's age and the length of interval. This allowed for consideration of the unique prediction from the GRTR screening measure in analyses that were comparably conservative to the age-standardized correlations reported for the earlier follow-up analyses. Preliminary analyses indicated that the GRTR was significantly, albeit not too highly, correlated with age at screening (i.e., $r = .42$, $p < .001$) but that it was uncorrelated with length of interval (i.e., $r = .13$, *ns*). Length of interval and age also were uncorrelated.

Results for the regression analyses and for the partial correlation analyses with the GRTR are shown in Tables 5 and 6. Overall, despite an average interval of 26 months and control for child age at baseline, the 20-item GRTR was a significant unique predictor of the criterion variable for at least 9 of 16 outcome variables in both the regression and partial correlation analyses. It is notable that for the regression analyses this included all four decoding measures, both language measures, and the three phonological processing measures from the CTOPP (i.e., PA, PM, and RN), even when there was significant prediction from one or both of the covariates. For the partial correlations results were quite similar, such that the correlation was significant for three of four decoding measures, both reading comprehension measures, both language measures, concepts of print, and PM. As was the case with the cross-sectional results, the absence of items on the screener that assessed vocabulary or other aspects of oral language notwithstanding, GRTR was a significant moderate predictor of both expressive and receptive vocabulary across the follow-up interval.

In the six instances where there was not a significant prediction in the regression analyses from the screener, age, length of interval, or both were significant and substantial predictors. Furthermore, for the easier blending, rhyming, and elision composite measures, as well as for letter-name and letter-sound knowledge, the follow-up sample showed significant negative skew and ceiling effects, a factor that likely attenuated the prediction from the screener (see Note 1). This supposition was supported

Table 5
Regression and Partial Correlation Analyses for *Get Ready to Read!* Screener (GRTR)
Predicting Emergent Literacy Outcome Measures Within Long-Term Follow-Up Sample,
Controlling for Age at Initial Assessment and Length of Interval

Follow-Up Measure	B	Standard Error	β/sr^2	R^2	GRTR Partial Corr.[a]
PPVT-R-SS (106)					
Age	−.04	.33	−.01/.00		
Interval	−.46	.27	−.16/.02		
GRTR	1.67	.45	.36***/.11	.14	.52*
EOWPVT-SS (105)					
Age	.42	.43	.10/.01		
Interval	−.70	.36	−.19/.03		
GRTR	1.32	.59	.23*/.05	.10	.50*
Blending composite (65)					
Age	.64	.14	.54***/.22		
Interval	.32	.12	.27*/.07		
GRTR	.20	.22	.11/.01	.39	.43
Elision composite (66)					
Age	.50	.16	.39**/.11		
Interval	.36	.14	.28*/.08		
GRTR	.22	.26	.11/.01	.25	.40
Rhyme composite (66)					
Age	.09	.08	.14/.01		
Interval	.27	.08	.38**/.14		
GRTR	.11	.14	.10/.01	.18	.28
Concepts About Print (69)					
Age	.22	.08	.36**/.09		
Interval	.13	.07	.19/.04		
GRTR	.19	.13	.18/.02	.24	.55*
CTOPP-PA (88)					
Age	−.41	.34	−.20/.04		
Interval	−.48	.29	−.17/.03		
GRTR	.95	.39	.26*/.06	.10	.47
CTOPP-RN (87)					
Age	−.18	.33	−.06/.00		
Interval	.21	.28	.08/.01		
GRTR	.75	.37	.23*/.05	.05	.06
CTOPP-PM (76)					
Age	−.91	.37	−.28*/.07		
Interval	.13	.31	.05/.00		
GRTR	1.21	.38	.36**/.12	.16	.57*

Note: Sample sizes at follow-up are noted parenthetically. R^2 reported for full model with GRTR and both covariates entered; models run with simultaneous entry. PPVT-R-SS = *Peabody Picture Vocabulary Test–Revised–Standard Score*; EOWPVT-SS = *Expressive One-Word Picture Vocabulary Test–Standard Score*; CTOPP = *Comprehensive Test of Phonological Processing*; PA = Phonological Awareness; RN = Rapid Naming; PM = Phonological Memory.
a. Partial correlation for GRTR controlling for age at screening and length of interval with long-term follow-up assessment.
*$p < .05$. **$p < .01$. ***$p < .001$.

by exploratory analyses that used the most difficult of the lower-level PA subtests, non-pictured syllable and phoneme-level elision, on which there were not significant skew or ceiling effects. Results of these analyses indicated that the screener was a significant predictor over and above both age and length of interval (i.e., $t = 2.35$, $p < .05$; $\beta = .28$). To further demonstrate the robustness of the findings, the partial correlation analyses were conducted restricting the length of the interval to between 20 and 30 months, capturing the majority of participants. These findings are in fact even stronger than those for the full sample, with the partial correlations between the GRTR and the criterion variables being significant with 11 of 16 outcome measures.

Table 6
Regression and Partial Correlation Analyses for *Get Ready to Read!* Screener (GRTR)
Predicting Letter and Reading Outcome Measures Within Long-Term Follow-Up Sample,
Controlling for Age at Initial Assessment and Length of Interval

Follow-Up Measure	B	Standard Error	β/sr^2	R^2	GRTR Partial Corr.[a]
Letter Name (69)					
Age	.46	.11	.51***/.18		
Interval	.41	.10	.41***/.16		
GRTR	−.03	.18	−.02/.00	.34	.18
Letter Sound (69)					
Age	.27	.04	.64***/.28		
Interval	.23	.04	.51***/.25		
GRTR	.04	.07	.06/.00	.60	.13
Word Identification (111)					
Age	.80	.32	.21*/.04		
Interval	1.80	.27	.50***/.24		
GRTR	1.11	.46	.20*/.03	.41	.53*
Word Attack (111)					
Age	.21	.17	.11/.01		
Interval	.76	.14	.43***/.19		
GRTR	.62	.24	.23*/.04	.30	.57*
Frequent Words (69)					
Age	.19	.10	.22/.03		
Interval	.52	.09	.54***/.29		
GRTR	.37	.17	.25*/.05	.41	.41
TOWRE-SS (64)					
Age	−1.01	.43	−.30*/.08		
Interval	−.34	.36	−.12/.01		
GRTR	1.05	.44	.30*/.08	.13	.66**
Passage Comp. (44)					
Age	.06	.34	.03/.00		
Interval	.84	.28	.43**/18		
GRTR	.50	.39	.19/.01	.24	.55*
GORT-4 ORQ (42)					
Age	−1.20	.61	−.29/.07		
Interval	−.55	.52	−.15/.02		
GRTR	2.59	.69	.55**/.27	.29	.68**

Note: Sample sizes at follow-up are noted parenthetically. R^2 reported for full model with GRTR and both covariates entered; models run with simultaneous entry. TOWRE-SS = *Test of Word Reading Efficiency–Standard Score*; Passage Comp. = *Passage Comprehension* subtest of the *Woodcock Johnson Test of Achievement*; GORT-4 = *Gray Oral Reading Test–4th Edition*; ORQ = Oral Reading Quotient.
a. Partial correlation for GRTR controlling for age at screening and length of interval with long-term follow-up assessment.
*$p < .05$. **$p < .01$. ***$p < .001$.

Discussion

Results from the concurrent and longitudinal analyses were consistent with those of previous studies in demonstrating that the GRTR screening measure is significantly related to more comprehensive measures of both emergent literacy foundational skills and conventional reading skills including both decoding and comprehension. Findings included significant concurrent and cross-time relations with these full-length assessments, including standardized measures of reading skills. Results indicated that even at intervals greater than 2 years, children's performance on the brief 20-item screening measure was predictive of later reading-related abilities.

The most notable aspect of this study is that the GRTR screener, which includes only a small number of items on each of three emergent literacy skill areas (i.e., PA, letter knowledge, print concepts), was both concurrently and predictively related to a wide array of reading-related measures. That is, not only did the GRTR predict scores on more comprehensive instruments for letter knowledge, PA, and print concepts, it also was significantly related to measures of oral language, word and nonword reading, reading comprehension, PM, and RN.

Previous research indicates that these skill areas are often moderately to highly related to each other, but they represent distinct abilities (e.g., Lonigan et al., 2000, 2007; Wagner et al., 1997). These results suggest that the screening tool not only is measuring three skill areas but also can be considered a proxy for preschool children's reading-related knowledge in general. These findings are consistent with the results of Molfese et al. (2004), who also found that the GRTR was concurrently related to numerous oral language and emergent literacy measures.

Although the results of this study are promising with respect to utility and predictive validity, the more stringent test of a reading-related screening measure is how well it is able to predict the categorical placement of participants or the need for diagnostic assessment (Lonigan, 2006). In other words, the most useful emergent literacy screening tools would be those that have sufficient positive and negative predictive power to help educators foreshadow which children are most at risk for reading failure once conventional instruction begins. Such information is of critical value when resources for early intervention are limited and should be given to the children most in need. As noted earlier, this study was intended only as a preliminary step toward addressing the issue of designating cut scores and using the GRTR as a formal screener for a longer measure that would fully ascertain children's risk status. However, these correlational and regression findings do indicate that even a measure as brief as the GRTR may be adequate to the task.

In fact, a recent study (Wilson & Lonigan, 2008b) supported the utility of the GRTR as a screening tool for a new, nationally standardized diagnostic measure of preschoolers' emergent literacy skills, the TOPEL (Lonigan et al., 2007), which contains three subscales representing PA, expressive vocabulary, and print knowledge. Of particular note is that several of the outcome measures (e.g., the eight PA tests) used in this article were used as prototypes for TOPEL items. Wilson and Lonigan (2008b) found that the GRTR was significantly correlated with the TOPEL composite Early Learning Index (ELI) and all three subtest scores both concurrently and longitudinally over a 3-month interval. These findings are consistent with the results of this study. Moreover, in a related study, Wilson and Lonigan (2008a) found that a cut score could be applied to the GRTR that yielded a high degree of accuracy and positive predictive power in predicting TOPEL ELI performance.

Few other brief instruments have been assessed in this manner, and success in this area has before now been reserved for larger screening batteries made up of multiple measures. For example, in an early longitudinal predictive study, Fletcher and Satz (1982) followed a cohort of students from kindergarten through sixth grade and found that the kindergarten assessments had relatively high utility for predicting sixth-grade categorical outcomes. Similarly, Uhry (1993) used a battery of kindergarten PA tasks to predict first-grade reading and correctly classified approximately 85% of readers. In one of the best prediction studies to date, O'Connor and Jenkins (1999) found that they could calibrate their cutoff scores across multiple measures to capture all the future reading-disabled children in their sample and have only a relatively small number of false positives (i.e., children thought to be at risk who developed into average-ability readers).

The development of a screening measure that could be used readily by teachers in preschools—and that has direct instructional relevance—is quite daunting because of the challenge of capturing within a brief measure the key skills of most predictive validity for later reading success, over and above the likely influence of demographic factors such as income, ethnicity, and cultural background. Moreover, to have generalizable utility, a screening measure has to maintain classification accuracy regardless of the specific instructional contexts in which children are taught currently or in the time period intervening between screening and criterion assessments. Despite these admitted challenges, the current results, together with those of Wilson and Lonigan (2008a, 2008b), support the viability of the much shorter and easier-to-administer GRTR as a screening tool and support the utility and benefit of, however briefly, measuring children's skills in more than one area (i.e., letter knowledge, PA). The inclusion of multiple skill areas and the careful selection of items to represent a developmental progression of difficulty may have helped mitigate against the common risk of obtaining floor effects. Single skill area instruments may be at greater risk of producing floor effects for preschool-age children and, thus, for having reduced predictive utility, especially within a high-risk sample. Research currently under way will be investigating the validity of the GRTR administered in preschool as a predictor of classification accuracy on measures of decoding and reading comprehension measured in kindergarten and first grade.

Although this study did not evaluate the utility and validity of a parent-administered form of the GRTR, pilot data indicate that parents can administer the scale with only minimal written training. The large numbers of parents who have accessed the GRTR via the NCLD Web site are also a testament to its approachability (K. G., NCLD, personal communication, December 11, 2006).

This ultimately may serve as a considerable advantage for this instrument over other measures that can be used only by trained educational personnel. Much more work is needed to investigate whether parent-administered measures (and teacher-administered measures) like this can retain the reliability and validity of the researcher-administered version.

A related aspect of the GRTR is that results from the NCLD-sponsored field trials and its quasi-experimental study of the GRTR as a teacher-administered tool (NCLD, 2003) indicate that the measure holds promise as an instructional aid within the preschool classroom. Teachers who participated in these trials reported learning a great deal about the skills needed for later reading success and having a better sense of what kind of activities to include within their curricula to help promote the development of these skills. NCLD also has developed a series of brief activity ideas for teachers and parents that are available on its Web site. Within NCLD's quasi-experimental study, those classrooms in which teachers received professional development on the GRTR, administered it in fall and spring in their classrooms, and used some of these recommended activities had students who showed more growth in skill than classrooms where students only were screened by outside assessors. In the absence of widespread availability of empirically supported preschool curricula that will reliably help preschool children grow in these abilities, a tool such as the GRTR screener may serve as a useful intermediary and adjunct device to promote greater understanding of and attention to important emergent literacy skills within the preschool classroom. NCLD continues to engage in large-scale training on the GRTR across the United States (NCLD, 2003).

This study had several limitations. First, because the participating students were recruited from three separate longitudinal studies, there was greater inconsistency in the measures administered at follow-up and in the length of follow-up intervals than would be desired. However, the benefit of this sampling strategy was that it substantially increased the ethnic and SES diversity of the baseline and thus the longitudinal samples. Second, the age range at baseline was greater than anticipated, and whereas the results appear to be robust to statistical controls and sampling restrictions related to age, it may be that this age range affected the findings. Future studies with narrower age bands or with sample sizes large enough to directly compare between age ranges would be valuable. Third, the sample size at the longitudinal follow-up was smaller than anticipated, although given the transitory lifestyle of many participating families, this was not entirely unexpected.

Whereas these factors suggest that some findings should be considered preliminary, the fact that many significant predictive relations were found despite the long intervals and small samples suggests that the findings are generally robust to these sampling issues. This is quite encouraging, given that longitudinal and intervention research with children from very low-income, more transient backgrounds, like many of this study's participants, is of significant importance because of their known elevated risk for reading difficulties. One other limitation was that little information was available with regard to the participants' exposure to reading instruction once they began elementary school. The children attended a number of public and private schools that likely varied widely in their reading curricula. However, the ability of the GRTR to predict outcomes up to 2 years after being administered attests to the robust predictive nature of the skills that the GRTR taps and to the stability of these skills, despite likely significant variability in the quantity and quality of instruction received by children. Furthermore, this very diversity of instruction received in kindergarten and first grade by the participants makes it unlikely that it was some common educational exposure during the follow-up period, rather than the GRTR and the skills it measures themselves, that was responsible for the significant longitudinal relations. Further research on the predictive power of this and other preschool screening measures could account for the diversity of educational exposure experienced by participating children. Further research also should expand on this study's efforts to include a diverse sample with respect to demographic variables by including children of Hispanic backgrounds and by more explicitly modeling the effect of such variables on children's assessment scores.

Note

1. Analyses with reflected and log-transformed total scores for the blending, elision, and rhyming composite scores and for Concepts About Print (CAP) revealed largely comparable results. *Get Ready to Read!* was a marginally significant predictor (i.e., $p < .10$) for Blending Composite and CAP but remained nonsignificant for the other outcome measures. Transformations were unsuccessful in correcting the overwhelming skew of the Letter Name and Letter Sound variables; therefore, these companion regression analyses were not conducted.

References

Badian, N. A. (1988). The prediction of good and poor reading before kindergarten entry: A nine-year follow-up. *Journal of Learning Disabilities*, *21*, 98–103.

Badian, N. A. (1994). Preschool prediction: Orthographic and phonological skills, and reading. *Annals of Dyslexia*, *44*, 3–25.

Badian, N. A. (1998). A validation of the role of preschool phonological and orthographic skills in the prediction of reading. *Journal of Learning Disabilities*, *31*, 472–481.

Berninger, V. W., Abbott, R. D., Vermeulen, K., Ogier, S., Brooksher, R., Zook, D., et al. (2002). Comparison of faster and slower responders to early intervention in reading: Differentiating features of their language profiles. *Learning Disabilities Quarterly*, *25*, 59–76.

Blatchford, P., Burke, J., Farquhar, C., & Plewis, I. (1987). Associations between pre-school reading related skills and later reading achievement. *British Educational Research Journal*, *13*, 15–23.

Bowey, J. A. (1995). Socioeconomic status differences in preschool phonological sensitivity and first-grade reading achievement. *Journal of Educational Psychology*, *87*, 476–487.

Bramlett, R. K., Rowell, R. K., & Mandenberg, K. (2000). Predicting first grade achievement from kindergarten screening measures: A comparison of child and family predictors. *Research in the Schools*, *7*, 1–9.

Catts, H. W., Fey, M. E., Zhang, X., & Tomblin, J. B. (1999). Language basis of reading and reading disabilities: Evidence from a longitudinal investigation. *Scientific Studies in Reading*, *3*, 331–361.

Chaney, C. (1998). Preschool language and metalinguistic skills are links to reading success. *Applied Psycholinguistics*, *19*, 433–446.

Chew, A. L., & Morris, J. D. (1989). Predicting later academic achievement from kindergarten scores on the Metropolitan Readiness Tests and the Lollipop Test. *Educational and Psychological Measurement*, *49*, 461–465.

Clay, M. M. (1979). *The early detection of reading difficulties* (3rd ed.). Portsmouth, NH: Heinemann.

Coyne, M. D., Kame'enui, E., Simmons, D. C., & Harn, B. A. (2004). Beginning reading intervention as inoculation or insulin: First-grade reading performance of strong responders to kindergarten instruction. *Journal of Learning Disabilities*, *37*, 90–104.

CTB/McGraw-Hill. (1990). *Developing Skills Checklist*. Monterey, CA: Author.

de Jong, P. F., & van der Leij, A. (2003). Developmental changes in the manifestation of a phonological deficit in dyslexic children learning to read a regular orthography. *Journal of Educational Psychology*, *95*, 22–40.

Duncan, J., & Rafter, E. M. (2005). Concurrent and predictive validity of the Phelps Kindergarten Readiness Scale–II. *Psychology in the Schools*, *42*, 355–359.

Dunn, L. M., & Dunn, L. M. (1981). *Peabody Picture Vocabulary Test–Revised*. Circle Pines, MN: American Guidance Service.

Fletcher, J. M., & Satz, P. (1982). Kindergarten prediction of reading achievement: A seven-year longitudinal follow-up. *Educational and Psychological Measurement*, *42*, 681–685.

Flynn, J., & Rahbar, M. H. (1998). Kindergarten screening for risk of reading failure. *Journal of Psychoeducational Assessment*, *16*, 15–35.

Fowler, M. G., & Cross, A. W. (1986). Preschool risk factors as predictors of early school performance. *Journal of Developmental & Behavioral Pediatrics*, *7*, 237–241.

Gardner, M. F. (1990). *Expressive One-Word Picture Vocabulary Test*. Novato, CA: Academic Therapy.

Havey, J. M., Story, N., & Buker, K. (2002). Convergent and concurrent validity of two measures of phonological processing. *Psychology in the Schools*, *39*, 507–514.

Invernizzi, M., Juel, C., Swank, L., & Meier, J. (2007). *Technical Manual for the Phonological Awareness Literacy Screening–Kindergarten*. Charlottesville: University of Virginia.

Invernizzi, M., Sullivan, A., Meier, J., & Swank, L. (2004). *Teacher's Manual for the Phonological Awareness Literacy Screening–Pre-Kindergarten*. Charlottesville: University of Virginia.

Juel, C. (1988). Learning to read and write: A longitudinal study of 54 children from first through fourth grades. *Journal of Educational Psychology*, *80*, 437–447.

La Paro, K. M., & Pianta, R. C. (2000). Predicting children's competence in the early school years: A meta-analytic review. *Review of Educational Research*, *70*, 443–484.

Lonigan, C. J. (2006). Development, assessment, and promotion of preliteracy skills. *Early Education and Development, 17,* 91–114.

Lonigan, C. J., Anthony, J. L., Phillips, B. M., Purpura, D. J., Wilson, S. B., & McQueen, J. (in press). The nature of preschool phonological processing abilities and their relations to vocabulary, general cognitive abilities, and print knowledge. *Journal of Educational Psychology*.

Lonigan, C. J., Bloomfield, B. G., Anthony, J. L., Bacon, K. D., Phillips, B. M., & Samwel, C. S. (1999). Relations among emergent literacy skills, behavior problems, and social competence in preschool children from low- and middle-income backgrounds. *Topics in Early Childhood Special Education*, *19*, 40–53.

Lonigan, C. J., Burgess, S. R., & Anthony, J. L. (2000). Development of emergent literacy and early reading skills in preschool children: Evidence from a latent variable longitudinal study. *Developmental Psychology*, *36*, 596–613.

Lonigan, C. J., Schatschneider, C., & Westberg, L. (in press). Results of the National Early Literacy Panel research synthesis: Identification of children's skills and abilities linked to later outcomes in reading, writing, and spelling. *Report of the National Early Literacy Panel*.

Lonigan, C. J., Wagner, R. K., Torgesen, J. K., & Rashotte, C. (2007). *The Test of Preschool Early Literacy*. Austin, TX: Pro-Ed.

May, D. C., & Kundert, D. K. (1997). School readiness practices and children at risk: Examining the issues. *Psychology in the Schools*, *34*, 73–84.

Molfese, V. J., Modglin, A. A., Beswick, J. L., Neamon, J. D., Berg, S. A., Berg, C. J., et al. (2006). Letter knowledge, phonological processing, and print knowledge: Development in nonreading preschool children. *Journal of Learning Disabilities*, *39*, 296–304.

Molfese, V. J., Molfese, D. L., Modglin, A. T., Walker, J., & Neamon, J. (2004). Screening early reading skills in preschool children: Get ready to read. *Journal of Psychoeducational Assessment*, *22*, 136–150.

Morrison, F. J., Griffith, E. M., & Alberts, D. M. (1997). Nature-nurture in the classroom: Entrance age, school readiness, and learning in children. *Developmental Psychology*, *33*, 254–262.

National Center for Learning Disabilities. (2003). *Evaluation findings from national demonstrations 2001–2003: Executive summary*. Retrieved September 19, 2004, from http://www.getreadytoread.org/pdf/NationalDemo.pdf

O'Connor, R. E., & Jenkins, J. R. (1999). Prediction of reading disabilities in kindergarten and first grade. *Scientific Studies of Reading*, *3*, 159–197.

Pae, H. K., Wise, J. C., Cirino, P. T., Sevcik, R. A., Lovett, M. A., Wolf, M., et al. (2005). The Woodcock Reading Mastery Test: Impact of normative changes. *Assessment*, *12*, 347–357.

Pianta, R. C., & McCoy, S. J. (1997). The first day of school: The predictive validity of early school screening. *Journal of Applied Developmental Psychology*, *18*, 1–22.

Reid, D. K., Hresko, W. P., & Hammill, D. D. (2001). *Test of Early Reading Ability–3rd Edition*. Austin, TX: Pro-Ed.

Schatschneider, C., Fletcher, J. M., Francis, D. J., Carlson, C., & Foorman, B. R. (2004). Kindergarten prediction of reading skills: A longitudinal comparative analysis. *Journal of Educational Psychology*, *96*, 265–282.

Shepard, L. A. (1997). Children not ready to learn? The invalidity of school readiness testing. *Psychology in the Schools*, *34*, 85–97.

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

Stipek, D. (2002). At what age should children enter kindergarten? A question for policy makers and parents. *Social Policy Report*, *16*(2), 3–16.

Storch, S. A., & Whitehurst, G. J. (2002). Oral language and code-related precursors to reading: Evidence from a longitudinal structural model. *Developmental Psychology*, *38*, 934–947.

Tabatchnick, B. G., & Fidell, L. S. (2001). *Understanding multivariate statistics* (4th ed.). Boston: Allyn & Bacon.

Texas Education Agency. (1999). *Texas Primary Reading Inventory*. Austin, TX: Author.

Torgesen, J. K. (2000). Individual differences in response to early interventions in reading: The lingering problem of treatment resisters. *Learning Disabilities Research and Practice*, *15*, 55–64.

Torgesen, J. K., & Bryant, B. R. (2004). *Test of Phonemic Awareness–2nd Edition*. Austin, TX: Pro-Ed.

Torgesen, J., K., Wagner, R. K., Herron, J., & Rashotte, C. (1998). *Prevention of reading problems using computer assisted instruction: A comparison of two methods*. Unpublished manuscript. Florida State University.

Torgesen, J. K., Wagner, R. K., & Rashotte, C. A. (1999). *Test of Word Reading Efficiency*. Austin, TX: Pro-Ed.

Tramontana, M. G., Hooper, S. R., & Selzer, S. C. (1988). Research on the preschool prediction of later academic achievement: A review. *Developmental Review*, *8*, 89–146.

Uhry, J. K (1993). Predicting low reading from phonological awareness and classroom print. *Educational Assessment, 1*, 349–368.

Wagner, R. K., Torgesen, J. K., & Rashotte, C. A. (1994). The development of reading-related phonological processing abilities: New evidence of bi-directional causality from a latent variable longitudinal study. *Developmental Psychology*, *30*, 73–87.

Wagner, R. K., Torgesen, J. K., & Rashotte, C. A. (1999). *Comprehensive Test of Phonological Processing*. Austin, TX: Pro-Ed.

Wagner, R. K., Torgesen, J. K., Rashotte, C. A., Hecht, S. A., Barker, T. A., Burgess, S. R., et al. (1997). Changing causal relations between phonological processing abilities and word-level reading as children develop from beginning to fluent readers: A five-year longitudinal study. *Developmental Psychology*, *33*, 468–479.

Whitehurst, G. J. (2001). The NCLD *Get Ready to Read!* screening tool technical report. Retrieved November 18, 2003, from http://www.getreadytoread.org/pdf/TechnicalReport.pdf

Wiederholt, J. L., & Bryant, B. R. (2001). *Gray Oral Reading Test–4th Edition*. Austin, TX: Pro-Ed.

Wilson, S. B., & Lonigan, C. J. (2008a). *A direct comparison of two emergent literacy screeners*. Manuscript in preparation.

Wilson, S. B., & Lonigan, C. J. (2008b). *Emergent literacy screeners for preschool children: An evaluation of* Get Ready to Read! *and individual growth and development indicators*. Manuscript in preparation.

Woodcock, R. W. (1987). *Woodcock Reading Mastery Test–Revised*. Circle Pines, MN: American Guidance Service.

Beth M. Phillips, PhD, is an assistant professor in the Department of Educational Psychology and Learning Systems at Florida State University and the Florida Center for Reading Research. Her current research interests include early literacy and behavioral development, preschool curriculum and instruction, and parental influences on learning.

Christopher J. Lonigan, PhD, is a professor of psychology at Florida State University and an associate director of the Florida Center for Reading Research. His areas of expertise include the development, assessment, and promotion of early literacy skills.

Marcy A. Wyatt is a research associate in the Department of Psychology at Florida State University. She directs longitudinal assessment projects related to early childhood literacy development.