National Center for Learning Disabilities
The power to hope, to learn, and to succeed

Technical Report on the Development of the NCLD Spanish-Language *Get Ready to Read!* Screening Tool

Prepared for the National Center for Learning Disabilities

Christopher J. Lonigan, Ph.D.
Florida State University and the Florida Center for Reading Research
December 2003

Background

Reading skills provide the foundation for children's academic success. Children who read well read more and, as a result, they acquire more knowledge in numerous domains. In contrast, children who lag behind in the development of reading skills receive less practice in reading than other children, miss opportunities to develop reading comprehension strategies, often encounter reading material that is too advanced for their skills, and may acquire negative attitudes about reading itself (Snow, Burns, & Griffin, 1998; Whitehurst & Lonigan, 2001).

Although the development of skilled reading occurs without significant problems for the majority of children, an estimated one in three children experiences significant difficulties in learning to read. Research has shown that there is strong continuity between the skills with which children enter school and their later academic performance. Children with limited reading-related skills rarely catch up to their peers. The impact of poor reading skills often resonates throughout a child's educational experience and into adulthood. For instance, Juel (1988) reported that the probability that children would remain poor readers at the end of the fourth grade if they were poor readers at the end of the first grade was .88. Children who are poor readers are frequently referred to special education classes (Lentz, 1988), and of those who experience the most serious reading problems (e.g., children scoring in the lowest 20th percentile), 10 to 15% drop out of high school, and only 2% complete a 4-year college program (NICHD, 2000).

During the past decade, a growing body of research evidence has highlighted the significance of the preschool period for the development of critically important early literacy skills (e.g., see Snow et al., 1998; Whitehurst & Lonigan, 2001). Data from these research studies indicate that oral language, phonological processing skills, and print knowledge are strongly predictive of

how well and how easily children will learn to read and write once they are exposed to formal reading instruction from kindergarten through the third grade. Oral language refers to the words in a child's vocabulary as well as his or her ability to use those words to convey meaning (i.e., syntactic and narrative skills). Phonological processing skills refer to children's developing sensitivity to the sound structure of their language (e.g., that words are made up of smaller sounds like syllables or phonemes). Print knowledge refers to a developing understanding about the nature and purpose of books and print (e.g., the sounds letters represent, directionality of print). Many children arrive at kindergarten with low levels of these skills, making it less likely that they will benefit from the instruction they will receive in the early elementary grades. Research also indicates that the level of emergent literacy skills a child has by the time he or she is four-years of age are likely to be strong predictors of his or her later reading skills. Differences between children with high versus low levels of emergent literacy skills tend to persist for children lagging behind in the development of these skills in the absence of interventions that directly target the skills (Lonigan, 2003).

Accurate assessment of the skills involved in the causal chain of learning to read and write is important because early identification can lead to focused early intervention. Data from a variety of sources indicate that the prevention of reading difficulties is likely more efficacious and cost - effective than remediation of reading difficulties. Children who have difficulty acquiring the alphabetic principle and who continue to experience problems with decoding miss the opportunity to develop the fluency required to become skilled readers. Early identification of areas in which children need targeted intervention is essential if the cycle of failure is to be broken.

The National Center for Learning Disabilities (NCLD), in collaboration with researchers involved in early literacy and literacy research, developed a screening tool to gauge preschool children's development in reading-related skills. This measure, the Get Ready to Read! screening tool (GRTR), was designed for English-speaking four-year-old children. GRTR is a 20-item multiple choice measure that provides information on children's early literacy skills, primarily in the areas of print knowledge (letter knowledge, print concepts) and phonological awareness. A complete description of the development and psychometric properties of the English-language GRTR screening tool is included in Whitehurst (2001; available online at http://www.getreadytoread.org). The English-language GRTR screening tool has acceptable

internal consistency reliability (α = .78), and it has good concurrent validity as measured by its correlations with a diagnostic measure of early literacy skills (the Developing Skills Checklist; CTB-McGraw-Hill, 1990; *r* = .69), a measure of receptive vocabulary (the Peabody Picture Vocabulary Test; Dunn & Dunn, 1981; *r* = .58), a measure of letter knowledge (*r* = .66), and a measure of phonological awareness (*r* = .58).

Development of the Spanish-Language GRTR Screening Tool

Children from families with limited English proficiency (i.e., children who are second language learners) are identified as a group who are at risk of reading difficulties (Snow et al., 1998). However, there are few measures that have been developed and validated for use with English language learners. Research suggests that the "code" components of early literacy skills transfer to reading in English (e.g., Comeau, Cormier, Grandmaison, & Lacroix, 1999). That is, children with strong phonological awareness and print knowledge skills in their home language tend to transfer those skills readily to English when reading instruction occurs in English. However, most early literacy assessment measures rely upon language to assess the literacy domain. A child asked to complete a phonological awareness or print knowledge task using words that he or she does not understand is likely to be at a severe disadvantage compared to a child who does understand the words used in the task. Consequently, low scores on a measure may reflect children's language skills rather than their ability levels in the domain being assessed. Measures designed for use and validated with children who are English language learners are needed to accurately assess key domains of emergent literacy.

Subsequent to the release of the English-language GRTR screening tool, NCLD began development of a Spanish-language version to meet the needs of Spanish-speaking four-year-old children. The goal of the development of the Spanish-language version of the GRTR was to create near parallel English and Spanish versions of the measure. That is, this development work was designed to produce a Spanish-language screening tool that assessed the same domains of children's early literacy skills (i.e., print knowledge and phonological awareness) as the English-language GRTR screening tool. The Spanish-language GRTR was an adaptation of the 20-items of the English-language GRTR. That is, Spanish-language items that had similar content to the items in the English-language GRTR measure were created by taking into account differences between Spanish and English where necessary. This adaptation resulted in a 20-item Spanish-

language measure that assessed print knowledge and phonological awareness in the same proportions and sequence as the 20-items of the English-language GRTR. Native speakers of Spanish reviewed these 20-items for accuracy and clarity, as well as to identify any potential source of bias in the items when used with a Spanish-speaking population.

Following the development of the 20-item Spanish-language version of the Get Ready to Read! screening tool, field-testing was conducted to determine the psychometric characteristics of the measure. The remainder of this report details the initial testing of and subsequent modifications to the Spanish-language GRTR.

Initial Field-Testing

In Spring 2002, the original 20-item Spanish-language GRTR (S-GRTR) was administered to 82 Spanish-speaking four- and five-year-old Latino preschool children attending Head Start centers in Los Angeles, California. These 82 children had a mean age of 58.82 months (*SD* = 4.72; range 49 to 67 months) and 45.7% were girls. The mothers of the majority of the children had immigrated to the United States from Mexico (i.e., > 50%). The mothers of the remainder of the children had either immigrated to the United States from Central America (i.e., > 20%) or been born in the United States. All children were bilingual but had English language skills ranging from limited to conversationally competent. Analyses of children's responses to the items on the S-GRTR revealed moderate internal consistency (α = .68) on the measure as a whole. This level of internal consistency was less than that produced by the English-language version (i.e., α = .78 for the entire development sample; α = .75 for the Head Start sample) of the measure (see Whitehurst, 2001). Inspection of the performance of individual items in the measure indicated that Item-17, an item assessing phonological awareness that required children to blend two words into a compound word was not performing adequately (i.e., item-total *r* = .00). A possible explanation for this result is that compound words are not as common in Spanish--particularly at the preschool vocabulary level--as they are in English. Several other items performed in the marginal range. Items 6, 10, and 15 had item-total correlations that were .15 or lower (see Table 1). To maintain the parallelism between the English- and Spanish-language versions of the GRTR, six potential replacement items within the domain of phonological awareness were generated for the original S-GRTR Item-17. Piloted testing of these items revealed that preschool children were capable of completing the items.

Expanded Field-Testing

In Spring 2003, initial trials using a 26-item version of the S-GRTR (all 20 original items and the 6 replacement phonological sensitivity items) were initiated with a group of Spanish-speaking preschool children attending Head Start centers in Los Angeles, California (see below for a description of the sample). Although the initial objective of this testing was to identify the best replacement item for the original Item-17, analyses of data from the first 20 - 30 children who completed the measure revealed that the pattern of results was not replicating the results from the initial field trial. That is, in this sample, original Item-17 was performing adequately; however, other original items were not performing adequately (i.e., several items had low item-total correlations). All replacement items had high item-total correlations, suggesting that they were measuring the skill-set that was measured by the S-GRTR as a whole. Consequently, the 26-item version of the S-GRTR was administered to the full sample of available Spanish-speaking preschool children to identify those items that best measured emergent literacy in Spanish-speaking preschool children and would provide the most internally consistent version of the S-GRTR that was closely parallel to the E-GRTR.

Expanded field-testing was conducted with 230 Spanish-speaking Latino children from Los Angeles, California. Eight children provided incomplete or unusable data. These children were either not available to complete all of the measures, refused to complete all of the measures, or gave clearly invalid responses (e.g., always selecting a response in a particular position). The final sample consisted of 222 children. These children had a mean age of 52.23 months (*SD* = 5.16; range: 42 to 62 months) and 51.4% were girls. Based on data provided from the 71% of the children's mothers (or maternal guardian) who agreed to complete a brief interview, the majority of the mothers of the children had immigrated from Mexico (67%). An additional 14% of the mothers had immigrated from Central America, and 9% of the mothers had been born in the United States. At the time of the assessment, the mothers who reported immigrating from Mexico or Central America had lived in the United States for an average of 17.7 years (*SD* = 8.17; range: 1 to 42 years). Virtually all of the children (97%) had been born in the United States. All children were bilingual but had English language skills that ranged from limited to conversationally competent.

In addition to completing the 26-item S-GRTR, all children completed the Spanish-language version of the Preschool Comprehensive Test of Phonological and Print Processing (Pre-

CTOPPP; Lonigan, Farver, & Eppe, 2002), a diagnostic measure of preschool children's emergent literacy that assesses skills in phonological awareness, phonological memory, and print knowledge domains. All children also completed the Preschool Language Scales (PLS; Zimmerman, Steiner, & Pond, 2002) in both Spanish and English to provide an estimate of children's overall oral language skills in both Spanish and English. Scores on the PLS revealed average verbal abilities in Spanish (mean standard score = 96.15, *SD* = 18.44) and below average verbal abilities in English (mean standard score = 78.52, *SD* = 17.04), consistent with the Spanish-language dominant designation of these bilingual children.

Final Item Selection for the S-GRTR Screening Tool

Analyses of the data from the expanded field-testing were directed at identifying the best set of 20-items for assessing emergent literacy skills in Spanish-speaking preschool populations. This 20-item set was determined by examining which of the 26 items performed most adequately in the sample of 222 four-year-olds who participated in the expanded field-testing. Indications that an item in a measure is performing adequately is provided by three sources of information:

- Relatively high item-total correlations (i.e., an item measures skills similar to those measured by the scale as a whole),
- The relative difficulty level of an item (i.e., neither too few nor too many children get the item correct, which allows scores on an item to discriminate between high achieving and low achieving children within a domain), and
- Evidence of validity (i.e., the items and the scale as a whole correlate with other measures designed to measure the construct of interest).

Item-total correlations for the original 20-items and the 6 replacement items are shown in Table 1 for the sample using in initial field-testing (Sample 1), the sample used in the expanded field-testing (Sample 2), and the combined samples where possible. As can be noted from the table, Item-1 and Item-6 performed inadequately in Sample 2 and in the combined samples (i.e., item-total correlations were near zero or negative for these two items). Additionally, Item-7 and Item-10 performed marginally in Sample 2 and in the combined samples. All other items had adequate item-total correlations (i.e., item-total $r > .20$).

Table 1

Item-Total Correlations for Screening Tool Items Within Item Sets and Samples

Item	Original 20 (Sample 1)	Original 20 (Sample 2)	Original 20 (Samples 1 & 2)	Alternative 20 (Sample 2)	All 26 Items (Sample 2)
1	.31	-.03	.08	---	-.03
2	.24	.35	.32	.35	.34
3	.29	.38	.35	.38	.39
4	.38	.40	.40	.40	.38
5	.33	.29	.30	.32	.31
6	.14	.05	.08	---	.04
7	.10	.13	.12	---	.14
8	.26	.35	.32	.35	.32
9	.21	.20	.20	.21	.23
10	.15	.14	.14	.16	.16
11	.37	.35	.35	.31	.31
12	.25	.32	.30	.31	.29
13	.35	.37	.36	.38	.36
14	.24	.19	.21	.18	.21
15	.25	.29	.28	.31	.31
16	.19	.31	.28	.26	.27
17	.00	.25	.17	.30	.28
18	.45	.31	.34	.39	.40
19	.26	.39	.34	.45	.46
20	.38	.23	.27	.26	.29
21	---	---	---	---	.27
22	---	---	---	---	.22
23	---	---	---	.39	.38
24	---	---	---	.37	.37
25	---	---	---	.36	.35
26	---	---	---	---	.23

Item difficulties for the original 20 items for Sample 1 and the combined samples, and for the 26 items for Sample 2 are shown in Table 2 as the percentage of children who correctly completed the item. Table 2 also lists the construct assessed by each of the items in the original 20-item and the 26-item versions of the measure. With the exception of Item-1, most items fell within the desired range of difficulty. The majority of items were completed correctly by between 30 and 70 percent of children.

Point-biserial correlations between items and scores on the subscales of the Pre-CTOPPP and PLS scores for Sample 2 are shown in Table 3. The empirical designation of items, based on the pattern of correlations between an item and the subscales of the Pre-CTOPPP, is shown in the rightmost column of Table 3. Item-1 had no significant correlation with an external measure. Item-7 had a relatively weak correlation with the external measure of print knowledge, and Item-22 had a moderate correlation with an external measure of phonological awareness. Items 6, 14, 20, 21, and 26 had moderate correlations with two external measures. The remaining items had at least one strong correlation with one of the external emergent literacy measures.

Overall, the pattern of item-total correlations, item-difficulties, and correlations with external measures indicated that Item-1, Item-6, and Item-7 were the items from the original 20-item set that performed most inadequately. Of the six replacement items, Item-21, Item-23, Item-24, and Item-25 performed well (i.e., adequate item-total correlations, adequate correlations with external measures of emergent literacy). Based on these findings, an alternative 20-item set was created. In this alternative 20-item set, original Item-1, original Item-6, and original item-7 were replaced by replacement Item-23, replacement Item-24, and replacement Item-25. As seen in Table 1, this alternative 20-item set produced adequate item-total correlations across all items (i.e., item-total *r*s .16 to .45).

Table 2

Item Difficulties for Spanish Screening Tool (% of Children with Correct Response on Item)

Item Number (Skill Assessed by Item)	Sample 1	Sample 2	Combined
1 (Print Concepts)	.41	.11	.19
2 (Letter/Print Recognition)	.55	.62	.60
3 (Letter/Print Recognition)	.44	.50	.48
4 (Letter/Print Recognition)	.57	.60	.60
5 (Print Concepts)	.54	.45	.47
6 (Letter Name Knowledge)	.37	.34	.35
7 (Letter Name Knowledge)	.27	.33	.32
8 (Letter Sound Knowledge)	.52	.48	.49
9 (Letter Sound Knowledge)	.43	.41	.41
10 (Letter Sound Knowledge)	.29	.28	.28
11 (Writing/Print Concepts)	.78	.66	.69
12 (Writing/Print Concepts)	.68	.58	.61
13 (Writing/Print Concepts)	.70	.68	.68
14 (Phonological Awareness - ISM)	.46	.40	.42
15 (Phonological Awareness - ISM)	.26	.32	.30
16 (Phonological Awareness - RY)	.50	.39	.42
17 (Phonological Awareness - BL)	.59	.72	.68
18 (Phonological Awareness - BL)	.66	.74	.72
19 (Phonological Awareness - BL)	.63	.73	.70
20 (Phonological Awareness - ELS)	.32	.36	.35
21 (Phonological Awareness - BL)	---	.35	---
22 (Phonological Awareness - BL)	---	.85	---
23 (Phonological Awareness - BL)	---	.79	---
24 (Phonological Awareness - BL)	---	.78	---
25 (Phonological Awareness - BL)	---	.82	---
26 (Phonological Awareness - ELS)	---	.36	---

Note. For Phonological Awareness Items, ISM = initial sound matching, RY = rhyming, BL = blending, ELS = elision.

Table 3

Point-Biserial Correlations of Screening Tool Items with Scores on Spanish-Language Preschool CTOPPP Subscales and Preschool Language Scales (Sample 2 only)

Item	Blending	Elision	Print	NWrd Rep	Span PLS	Eng PLS	Empirical Definition
1	-.10	-.03	-.00	-.05	-.03	.01	None
2	.21**	.25***	.37***	.16*	.26***	.31***	EL/Print
3	.15*	.21***	.41***	.12	.14*	.26***	EL/Print
4	.20**	.27***	.39***	.19**	.14*	.14*	EL/Print
5	.27***	.29***	.34***	.22***	.19**	.16*	EL
6	.03	.14*	.23***	-.02	.10	-.12	Print
7	.00	.02	.16*	.04	.05	-.07	Print
8	.23***	.22***	.37***	.14**	.08	.32***	EL/Print
9	.09	.14*	.33***	.13*	.03	.07	Print
10	.09	.12	.27***	.04	.02	.08	Print
11	.21**	.17*	.30***	.21**	.18**	.26***	EL/Print
12	.20**	.25***	.26***	.19**	.06	.29***	EL
13	.17**	.14*	.35***	.23***	.21**	.16*	EL/Print
14	.05	.18**	.19**	.04	-.01	-.02	Print
15	.12	.18**	.27***	.20**	.18**	.19**	EL/Print
16	.04	.15*	.23***	.14*	.06	.08	EL/Print
17	.09	.21**	.17*	.20**	.14*	.15*	EL
18	.11	.24***	.23***	.17*	.09	.13*	EL
19	.18**	.23***	.34***	.30***	.15*	.23***	EL
20	.07	.17**	.17**	.11	.17**	.09	EL
21	.02	.13	.22***	.05	.20**	.02	Print
22	-.02	.14*	.13	-.03	.10	.09	PA
23	.15*	.23***	.27***	.13	.19**	.13	EL
24	.30***	.22***	.29***	.28***	.24***	.18**	EL
25	.25***	.25***	.23***	.22***	.16*	.07	EL
26	.00	.12	.19**	.12	.14**	-.08	Print

Note. N = 222. Pre-CTOPPP = Preschool Comprehensive Test of Phonological and Print Processing; NWrd Rep = NonWord Repetition; EL = emergent literacy; PA = phonological awareness. $^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$.

The overall internal consistency reliabilities (alphas) for both original and alternative item sets for both samples are shown in Table 4. As noted previously, the original 20-item set produced marginal internal consistency in Sample 1, Sample 2, and the combined samples. The alternative 20-item set produced a higher internal consistency reliability that was in the acceptable range (i.e, .70 or above); in fact, this alternative 20-item set producing an alpha that was higher than the alpha of the full 26-items. As a check on the reproducibility of these results, Sample 2 was split randomly into two subsamples of 111 children each. For the original 20-items (αs = .70 and .69) and the alternative 20-item set (αs = .75 and .77), the results of the two subsamples replicated the findings from the full sample.

Table 4

Internal Consistency Reliabilities for Screening Tool Using Different Item Sets

Item Set and Sample	Alpha	*N*
Original 20 Items (Sample 1)	.68	82
Original 20 Items (Sample 2)	.70	222
Original 20 Items (Sample 1 and 2 Combined)	.69	304
Full 26-Item Set (Sample 2)	.75	222
Alternative 20-Item Set (Sample 2)	.76	222

Correlations between total scale scores of the original and alternative versions of the S-GRTR and the external measures of emergent literacy skills and language are shown in Table 5. Given the high degree of overlap in the item content of the two versions of the S-GRTR, it would be expected that (a) the versions would be highly correlated and (b) there would be a similar pattern of correlations with the external measures across the two versions. The results shown in Table 5 confirm these expectations. Notably, both versions of the S-GRTR yielded significant and substantial correlations with the Elision (phonological awareness) and Print Knowledge subscales of the Spanish Pre-CTOPPP. The pattern of correlations between the S-GRTR and the

Spanish Pre-CTOPPP change little across versions, but to the extent that there are some minor differences in the correlations, the pattern reflects the change in content between the replaced and the replacement items. That is, original items 1, 6, and 7 all reflected print knowledge whereas replacement items 23, 24, and 25 all reflected phonological awareness; consequently, the alternative item set has stronger correlations with external measures of phonological awareness and the original item set has a stronger correlation with the external measure of print knowledge (although these differences do not approach statistical significance).

Table 5

Correlations of Item Sets with Scores on Spanish-Language Preschool CTOPPP Subscales and Preschool Language Scales (Sample 2 only)

	S-GRTR Item Sets	
	Original 20-Items	Alternative 20-Items
Alternative Item Set	.96***	---
Pre-CTOPPP Blending	.32***	.37***
Pre-CTOPPP Elision	.47***	.48***
Pre-CTOPPP Print	.70***	.68***
Pre-CTOPPP NonWord Repetition	.36***	.40***
Spanish PLS Standard Score	.29***	.31***
English PLS Standard Score	.37***	.39***

Note. N = 222. Pre-CTOPPP = Preschool Comprehensive Test of Phonological and Print Processing; PLS = Preschool Language Scale. ***$p < .001$.

Both the significance and relative size of the correlations between the two versions of the S-GRTR and the phonological awareness and print knowledge subscales of the Spanish Pre-CTOPPP were maintained even after controlling for the degree of variance shared by S-GRTR scores and Spanish Pre-CTOPPP scores with chronological age, Spanish PLS scores, and English PLS scores. Partial correlations controlling for age and PLS scores for the original 20-item and alternative 20-item versions, respectively, were .15 and .20 for the blending subscale and .33 and

.34 for the elision subscale of the Spanish Pre-CTOPPP. Partial correlations controlling for age and PLS scores for the original 20-item and alternative 20-item versions, respectively, were .63 and .59 for the print awareness subscale of the Spanish Pre-CTOPPP. Therefore, the size of the validity coefficients for the two different versions of the S-GRTR cannot be attributed to factors such as oral language proficiency or chronological age. Note that examining the relation between S-GRTR and Pre-CTOPPP scores while controlling for chronological age is a particularly stringent test of the validity of S-GRTR because both phonological awareness and print knowledge are in a period of rapid growth during a child's fourth year; hence, it would be expected that scores on measure of these constructs would increase with age even within a relatively narrow one-year interval. These results convincingly indicate that S-GRTR measures reliable variance related to emergent literacy skills in the domains that it is intended to measure.

Descriptive statistics for the two versions of S-GRTR in the two samples are shown in Table 6. Each version yielded a full range of scores. Inspection of the properties of the resultant distribution of scores indicated that both versions produced distributions that were approximately normal, with the alternative 20-item version producing somewhat better distributions than the original 20-item set.

Table 6

Descriptive Statistics of Total Scales for Different Screening Tool Item Sets

	Min.	Max.	Mean	*SD*	Median
Original 20 Items (Sample 1)	3	20	9.96	3.61	9.0
Original 20 Items (Sample 2)	1	19	9.69	3.62	10.0
Alternative 20-Item Set (Sample 2)	2	20	11.30	3.96	11.5

Girls scored approximately one point higher than did boys on both versions of the S-GRTR. Given the relatively large size of the sample, this small difference between boys and girls achieved statistical significance, $F(1, 220) > 6.13$, $p < .05$. Additionally, there was a small but significant age trend on both versions of the S-GRTR ($r = .31$ between chronological age and scores on the alternative 20-item version). This effect of age on scores on the alternative 20-item version is shown in Figure 1. Similar to the results for the E-GRTR (see Whitehurst, 2001), the

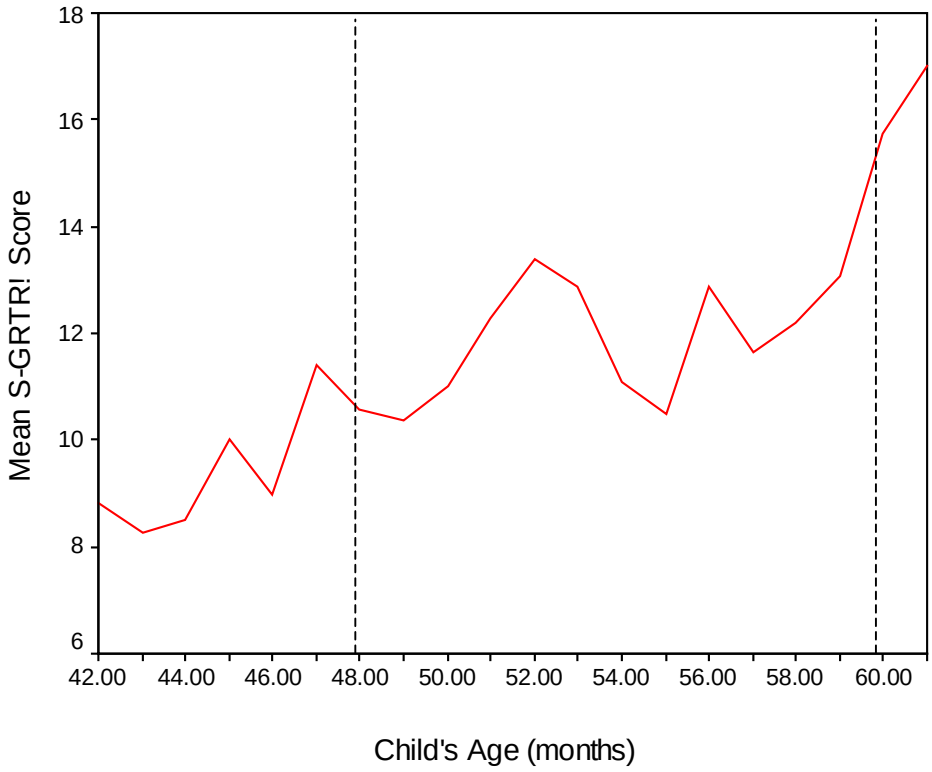

Figure 1. Scores on the alternative 20-item version of the Spanish-language GRTR! Screening Tool as a function of children's ages.

effect of age was due mainly to significantly higher scores for children older than 5-years and significantly lower scores for children younger than 4-years. Overall, scores on the S-GRTR increased by two-tenths of a point for each month of children's ages. Therefore, the effect of age across the period of 48- to 60-months equaled 1.2 points on the S-GRTR, an effect similar in size to the age trend from 48- to 60-months on the E-GRTR (see Whitehurst, 2001). As with the E-GRTR, these results indicate that use of the S-GRTR with children who are younger than 48-months or older than 60-months will require that age be taken into account when interpreting scores (e.g., a three-year-old child scoring significantly below the average range may still be exhibiting age-appropriate emergent literacy skills).

Given the moderate internal consistency reliability for either version of the scale, the pattern of item-level correlations with the external measures, and the fact that the GRTR was constructed to assess multiple domains of emergent literacy (i.e., phonological awareness, print knowledge), it is likely that the scale is multi-factorial. To test this possibility, an exploratory factor analysis was conducted with the 20 items from the alternative version of the scale. Initial

statistics indicated that there were between 1 and 7 factors represented in the items. Examination of solutions for different 2-, 3-, 4-, and 5-factor models revealed that no solution was interpretable along the content of the items (i.e., all phonological awareness items did not load on the same factor), most solutions had significant cross-loadings (i.e., items had significant factor loadings on two or more factors), and most solutions included nonsubstantial loadings for some items. Consequently, there is no evidence that a coherent multi-factorial organizational scheme (e.g., phonological awareness, print knowledge) can fit these data.

Summary

Overall, these results indicate that the S-GRTR provides a reliable and valid measure of emergent literacy skills for second language learners whose home-language is Spanish. Whereas the validity of the original 20-items is similar to the alternative 20-item version, the internal consistency reliability for the original 20-item scale is below the generally accepted minimum value of .70. Replacing three items improves the overall internal consistency of the scale. Using randomly selected subsamples of the children, these reliability estimates were replicated, suggesting that the increase in reliability was not entirely the result of empirical optimization. Whereas further optimization of scale internal consistency from the 26-item pool would have been possible, such procedures would have significantly reduced the parallelism between Spanish and English versions of GRTR (i.e., a much different balance of print versus phonological awareness items between Spanish and English versions).

It is worth noting that the reliability of the alternative 20-item version equals or exceeds the reliability reported for the English-language version of the GRTR screening tool (E-GRTR), and its reliability is substantially better than the reliability reported for the Hispanic subsample in the development sample for the E-GRTR (i.e., for the English-language version, $\alpha = .55$ for the Hispanic sample; see Whitehurst, 2001). Similarly, the validity coefficients for the S-GRTR are similar to those reported for the E-GRTR, particularly for print knowledge.

These results are significant for at least two reasons: First, the development sample for the E-GRTR explicitly included children from a diverse set of socioeconomic backgrounds (e.g., Head Start, fee-for-service preschools serving middle-income families). This inclusion of children from a wide range of backgrounds likely increased the range of scores obtained on the E-GRTR. By necessity, the sample used to develop and validate the S-GRTR included only children from

families of lower socioeconomic backgrounds. Given that children from families of lower socioeconomic backgrounds are often identified as at-risk for reading problems, the range of scores obtained was likely more restricted than the range of scores obtained in the development sample for the E-GRTR. Hence, it might be expected that range-restriction would negatively influence the psychometric outcomes of the S-GRTR, resulting in lower estimates of internal consistency reliability and validity. Second, the psychometric data reported for E-GRTR came from the development sample in which a pool of 60 items was reduced to 20 items based on optimization of internal consistency and correlations with external measures. Because internal consistency and validity coefficients were used in item selection for all items, thus optimizing these two psychometric metrics, it might be expected that both reliability and validity coefficients would be substantially higher for the E-GRTR than the S-GRTR.

The average performance of children is similar on the E-GRTR and the S-GRTR. Unlike the development sample for E-GRTR in which Hispanic children scored significantly lower than other children (i.e., mean score of 6.86; see Whitehurst, 2001), these results suggest that S-GRTR provides a relatively unbiased means to assess Spanish-speaking children's emergent literacy skills. Consequently, the same interpretive scale for scores developed for the E-GRTR should be used for the S-GRTR for Spanish-speaking four-year-old children (see Table 7).

Table 7

Score Interpretation Key for Spanish Get Ready to Read Screening Tool

Score on S-GRTR	Interpretation of Score on S-GRTR
0 - 6	Very Weak Skills
6 - 9	Weak Skills
9 - 12	Average Skills
12 - 16	Strong Skills
16 - 20	Very Strong Skills

In some settings, the GRTR will likely be used with preschool-age children generally, some of whom may be younger than 48-months and some of whom may be older than 60-months. Whereas the results of this field testing indicates that the screening tool can be used with these children, the interpretive scale shown in Table 7 cannot be used with children who are younger than 48-months or with children who are older than 60-months. That is, a three-year-old child would be expected to obtain lower scores than the average four-year-old and a five-year-old would be expected to obtain higher scores than the average four-year-old. Although the score would represent the child's emergent literacy skills and could be used to track his or her progress, the meaning of the score relative to his or her same-age peers is not known, and therefore, a interpretation of above, below, or at average cannot be made. Normative information for children younger than 48-months and older than 60-months should be obtained if an interpretative of the screening tool score is desired for these ages.

As noted by Whitehurst (2001) in the technical report for the English-language version of the Get Ready to Read! screening tool, where a child is regarding her or his preschool education is likely an important consideration in interpreting scores on the measure. That is, a child who scores in the average range on the measure at the beginning of the preschool year is likely in better shape for being ready to learn to read than a child who scores in average range at the end of the preschool year (i.e., after 9 months of educationally meaningful activities related to pre-reading skills). Hence, interpretation of scores on the S-GRTR should take into account where a child is in the academic year (i.e., no preschool versus a full year of preschool). The children who participated in this evaluation of the S-GRTR had already participated in over half a year of preschool education. Were they to have been assessed at the start of the preschool year, it is likely that their scores would have been lower.

It is not clear why the results of the item analysis differed between the sample used in the initial field-testing and the sample used in the expanded field-testing. One possibility is that a significant percentage of the sample used in the expanded field-testing was attending Head Start centers with a specific literacy focus. This exposure may have increased the likelihood that children had a better understanding of the tasks presented to them on the S-GRTR. Whereas this is a nominally viable explanation, the fact that the two groups of children scored similarly on the original 20-item set suggests that general emergent literacy content knowledge was not responsible for the difference. However, it is possible that children in Sample 2, who likely had

more exposure to phonological awareness activities than the children in Sample 1, were better able to perform a complex blending task with compound words.

A different field trial of the original 20-item S-GRTR screening tool with a group of 182 Spanish-speaking children from Georgia and Arizona produced results similar to those obtained with this sample. In the Georgia and Arizona sample, the internal consistency reliability of the scale was .71 and the mean score was 9.55. These results are almost identical to the results reported for this sample (see Table 4 and Table 6) for the original 20-item version of the S-GRTR in the samples used in the initial and expanded field-testing. The comparability of results across these different samples indicates that the psychometric properties of the S-GRTR are stable. Because the field-testing conducted in Georgia and Arizona did not use the expanded item pool (i.e., the alternative 20-item set), it is not known if the psychometric properties of the alternative version of the S-GRTR also would have replicated across the different samples; however, given the high degree of overlap in items and the close predictive correspondence between the original and alternative version of the measure, it is likely that the alternative version would have performed similarly in the Georgia and Arizona samples as it did in the Los Angeles sample.

Whereas the data provided both in the technical report for the English-language version of the GRTR and in this report provide substantial evidence of the concurrent validity of the GRTR screening tool in both English and Spanish, ultimately the value of the tool will be determined by its predictive validity. That is, the significant overlap of scores on the GRTR screening tool with scores on more comprehensive diagnostic tests, which are known to be predictive of later reading achievement, suggest that scores on the GRTR screening tool will also be predictive of later reading achievement. However, it is advisable to conduct evaluations of the predictive utility of the GRTR screening tool to enable direct empirical demonstrations of a predictive relationship between scores on the GRTR screening tool and later reading outcomes.

Finally, because the GRTR screening tool was constructed to assess multiple domains of emergent literacy (e.g., phonological awareness, print knowledge), it is tempting to compute separate subscale scores from the measure for each of these domains. This is inadvisable for at least two reasons. First, as demonstrated by the analyses of data generated in this sample, no coherent multi-factor solution fits the data. Hence, whereas the conceptual basis of including

specific items to assess different and distinct domains of emergent literacy guided the development of the screening tool, the empirical reality of children's responses to the screener's 20 items does not map directly to the conceptual model (see also Table 3). Second, because there are only 20 items in the measure, the number of items that target a specific domain are limited. Such a situation does not allow an adequate sampling of children's performance within a domain to allow valid statements about his or her skills related to a specific domain to be made. Diagnostic measures include a sufficient number of items for each content domain to adequately sample children's skills in an area. If the purpose of an assessment is specificity, only measures that are adequate to the task should be employed. Low scores on a screening measure should lead to additional assessment using diagnostic measures to uncover specific patterns of strengths and weaknesses.

References

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

Comeau, L., Cormier, P., Grandmaison, E., & Lacroix, D. (1999). A longitudinal study of phonological processing skills in children learning to read in a second language. *Journal of Educational Psychology, 91*, 29-43.

CTB (1990). *Comprehensive Test of Basic Skills.* Monterey, CA: CTB/McGraw Hill.

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.

Lentz, F. E. (1988). Effective reading interventions in the regular classroom. In J. L. Graden, J. E. Zins, & M. J. Curtis (Eds.), *Alternative educational delivery systems: Enhancing instructional options for all students*. Washington DC: National Association of School Psychologists.

Lonigan, C. J. (2003). Development and promotion of emergent literacy skills in preschool children at-risk of reading difficulties. In B. Foorman (Ed.), *Preventing and Remediating Reading Difficulties: Bringing Science to Scale* (pp. 23 - 50). Timonium, MD: York Press.

Lonigan, C. J., Farver, J. M., & Eppe, S. (2002). *Preschool Comprehensive Test of Phonological and Print Processing: Spanish-Language Version*. Tallahassee FL: Author.

NICHD 2000. *Report of the National Reading Panel: Teaching children to read.* Washington, DC: U.S. Department of Health and Human Services.

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

Whitehurst, G. J. (2001). *The NCLD Get Ready to Read! Screening Tool technical report.* Stony Brook, NY: Applied Research Partners (available on-line as http://www.getreadytoread.org/images/GRTR_%20Screen_Tech.pdf).

Whitehurst, G.J., & Lonigan, C.J. (2001). Emergent Literacy: Development from Prereaders to Readers. In S. B. Neuman & D. K. Dickenson (Eds.), *Handbook of Early Literacy Research* (pp. 11 - 29). New York: Guilford Press.

Zimmerman, I., Steiner, V., & Pond, R. (2002). *Preschool Language Scale-Fourth Edition.* San Antonio, TX: The Psychological Corporation.