pals™ español

Phonological Awareness Literacy Screening

Kindergarten

Technical Reference

Karen Ford • Marcia Invernizzi
University of Virginia • Curry School of Education

www.palsmarketplace.com

Illustrations: Charles Peale
Graphic Design: Branner Graphic Design

Printed in the United States of America

pals™ español

Phonological Awareness Literacy Screening

K Technical Reference

Karen Ford • Marcia Invernizzi
University of Virginia • Curry School of Education

Research for the development of PALS español was supported by the Institute of Education Sciences, U.S. Department of Education, through Grant R305A090015 to the University of Virginia.

Acknowledgments

We would like to thank the Institute of Education Sciences of the U.S. Department of Education, which provided funding for the development of PALS español K through a grant to the University of Virginia. We would also like to thank Dr. Xitao Fan at the University of Macau and Dr. Francis Huang at the University of Missouri for their contributions to the technical adequacy of PALS español K. Thanks go also to Dr. Igone Arteagoitia at the Center for Applied Linguistics for her advice on linguistic issues related to item development. Finally, we would like to thank our talented team of assessors, who have administered PALS español K to hundreds of children across the U.S. over the past four years, as well as the many teachers and children across the country who graciously agreed to participate in the pilots and field tests.

Section I

5 Introduction

5 Purpose and Uses
5 Overview
6 PALS español K Admistration and Scoring

Section II

7 Item Development and Field Testing

7 Preliminary Test Development and Initial Piloting
7 Pilot Testing
8 Field Testing

Section III

11 Establishing Summed Score Criteria and Benchmarks

Section IV

13 Technical Adequacy

13 Reliability
15 Validity

Section V

20 Summary

Section VI

21 References

Section I

Introduction

Purpose and Uses

The primary purpose of PALS español K is to identify students who perform below grade-level expectations in foundational early literacy skills and therefore may be at risk for developing reading difficulties or delays. PALS español K also serves as a diagnostic tool that provides valuable information that helps teachers target instruction to each student's individual literacy needs. PALS español K has demonstrated good evidence of reliability and validity; however, like any other assessment tool, it is just one means of measuring a student's overall literacy competence. Teachers should use PALS español K assessment data along with other sources of information, including additional assessment data, parent information, and their own judgment, in making decisions about students' literacy instruction.

Overview

Research has demonstrated the importance of phonological awareness (Branum-Martin et al., 2006; Bravo, 1995; Bravo, Villalón, & Orellana, 2006; Herrera & Defior, 2005; Jiménez & Ortiz, 2000; Manis, Lindsey, & Bailey, 2004), alphabet knowledge (Bravo et al., 2006; Gómez, 2008; Manis et al., 2004), and orthographic knowledge (Davies, Cuetos, & Glez-Seijas, 2007; Escribano, 2007; Escribano, Elosúa, Gómez-Veiga, & García-Madruga, 2013; Goswami, 2010) in predicting later reading achievement in Spanish among both monolingual and bilingual Spanish-speaking children.

Phonological awareness refers to the ability to attend to and reflect on the sounds of oral language apart from meaning and includes skills ranging from being able to recognize words that rhyme to the ability to segment individual phonemes. *Alphabet knowledge* includes both alphabet recognition (i.e., the ability to name the letters of the alphabet) and letter sound knowledge (i.e., the ability to produce the sounds made by letters). *Orthographic knowledge* is knowledge about words in their written form, including the application of grapheme-phoneme correspondences in spelling, an understanding of allowable word patterns in a language (e.g., *qu* is an acceptable letter combination in Spanish, but *qx* is not), and an awareness of morphological relationships between words (e.g., *joven*, *juventud*).

In a study that explored the importance of phonological awareness in Spanish, Branum-Martin and colleagues (2006) found that phonological tasks consisting of blending pseudowords, segmenting words, and phoneme elision, administered in Spanish in kindergarten, were significantly correlated with word reading in Spanish among 812 students in 74 transitional bilingual programs in California and Texas. In another study that followed 227 children from first to fourth grade in Chile, Bravo and his colleagues (2006) found that the skills that best predicted reading achievement at the end of first grade were letter naming and phonemic awareness, specifically phoneme segmentation and identification of initial phonemes. Carrillo (1994) examined the relationship between different levels of phonological awareness skills (i.e., less to more challenging) and learning to read in Spanish in a study with 120 kindergarten and first grade children in Murcia, Spain. She found that although the less demanding phonological awareness skills (e.g., sensitivity to rhyme and alliteration) were no longer correlated to reading in first grade, phonological awareness tasks at all levels, including sensitivity to rhyme, sensitivity to alliteration, and

Table 1 PALS español K Tasks

Tasks	Domains		
	Phonological Awareness	Alphabet Knowledge	Orthographic Knowledge
Rhyme Awareness	✓		
Beginning Sound Awareness	✓		
Alphabet/Digraph Recognition		✓	
Letter Sound Knowledge		✓	✓
Spelling	✓	✓	✓
Concept of Word	✓	✓	✓

phoneme segmentation and isolation, were significantly correlated to reading in kindergarten. Finally, in multiple studies conducted with elementary-age children in Spain, Escribano and colleagues (2007, 2013) found that orthographic processing is one of the strongest predictors of reading comprehension in Spanish. In fact, orthographic knowledge and reading speed were found to be much more closely related to reading comprehension in Spanish than was reading accuracy.

As shown in Table 1, PALS español K tasks are designed to measure students' progress toward developing the foundational skills in phonological awareness, alphabet knowledge, and orthographic knowledge that research has shown to be associated with reading achievement in Spanish.

PALS español K Administration and Scoring

PALS español K tasks are administered by teachers in a classroom setting. Tasks do not have a time limit and are administered individually, with the exception of the Group Rhyme Awareness, Beginning Sound Awareness, and Spelling tasks, which are administered in small groups of five or fewer students. Students receive a score on each PALS español K task, as well as an overall score, called a Summed Score, which is calculated by adding together scores on selected PALS español subtasks. Criterion scores, referred to as benchmarks, representing developmental expectations for fall and spring are provided for each task so that teachers can measure individual students' progress toward developing basic foundational literacy skills and can design instruction accordingly. Benchmarks are also provided for the PALS español K Summed Score for fall and spring. The Summed Score benchmark is used to identify students who may be at risk for developing reading difficulties.

Detailed procedures for administering and scoring PALS español K can be found in the Administration and Scoring Guide. A description of how PALS español K benchmarks were established can be found in Section III of this Technical Reference.

Section II

Item Development and Field Testing

Preliminary Test Development and Initial Piloting

In fall 2004, our research team began to explore the feasibility of developing an early literacy assessment in Spanish. After an extensive review of the research on literacy acquisition in Spanish and of current practices in Spanish literacy instruction, we concluded that there is scientific evidence supporting the existence of a systematic, developmental progression of skills in Spanish reading, writing, and spelling (e.g., Cuetos & Suárez-Coalla, 2009; Ferreiro, 1991; Ferreiro & Teberosky, 1982; Hachén, 2002; Manrique & Signorini, 1998; Mathes, Pollard-Durodola, Cárdenas-Hagan, Linan-Thompson, & Vaughn, 2007; Pollard-Durodola & Simmons, 2009), similar to the progression of skills that forms the basis for literacy assessment in English. Such a progression suggests that the same types of tasks that have proven successful in measuring English literacy development would also be appropriate to use in Spanish. Clearly, however, some tasks would need to be altered to some degree because of phonological and orthographic differences between the two languages. More importantly, individual items within tasks would need to be designed to reflect the specific skills associated with Spanish literacy development at each level.

Our first step in test and item development was to conceptualize the tasks and create an initial item pool. To test our basic construct, in spring of 2005 we piloted a draft of selected tasks with 125 children in Grades K–3 in a large metropolitan area in the northeastern United States. Participants spoke Spanish as their first language and were enrolled in bilingual education or dual language programs. All assessments were administered by classroom teachers in classroom settings.

Data from this preliminary pilot were analyzed, and results were used to make adjustments to the conceptualization of PALS español. From fall 2005 until spring 2008, we continued to make revisions to items and test them informally with students in Grades K–3. During this time, teachers in 19 schools in four metropolitan school districts in the northeastern and western regions of the U.S. administered PALS español K and 1–3 tasks to 2,083 students in Grades K–3. Many of these students were tested multiple times over multiple testing periods (i.e., fall and spring) and multiple years (i.e., 2005–06, 2006–07, and 2007–08 academic years). All participants were native Spanish speakers enrolled in bilingual or dual language programs.

Through these initial development efforts, we were able to explore the viability of our developmental construct for Spanish and gain valuable information on how the tasks and items on the Spanish instruments should be constructed. We then used that information to refine the existing tasks and items and create larger item pools in preparation for more formal pilot testing, which began in fall 2009.

Pilot Testing

During the 2009–2010 and 2010–2011 academic years, we conducted pilot tests of all PALS español K subtasks. For each task (except Letter Sounds and Alphabet and Digraph Recognition, which have a finite number of possibilities), we created sufficient items to support the production of two parallel forms of the assessment, as well as an additional 25% of items to allow for replacing any psychometrically poorly performing items that might be revealed in the empirical item trial process. For example, we

planned to have 10 items on the final Beginning Sound Awareness task. The item pool that we developed for piloting this task had 26 items, with 7 unique items allocated into each of the two parallel forms (7×2=14), and with 3 items that would be shared between the two forms as an equating block to ensure that scores from the parallel forms can be interpreted interchangeably. The remaining 9 items were reserve items to replace poorly performing items in the two parallel forms. During the piloting, each item was administered to a minimum of 200 kindergartners by assessors trained by the research team.

Assessors underwent rigorous training that involved explanation of the theoretical foundations of the assessment and instruction and practice in administering each PALS español task (subtask). We conducted a total of eight training sessions of approximately two hours each for a total of 16 hours. After each session, assessors were instructed to practice administering the tasks. They were then required to administer the tasks to each other while being evaluated by the research team. Each assessor was required to demonstrate that he/she could administer the assessment according to written protocol prior to conducting any testing in the field. Once in the field, assessors were observed to ensure valid administration procedures and reliable scoring. During training and observations in the field, assessors were scored on a number of reliability and validity dimensions using a rubric developed by the research team. Each assessor was observed until he/she met reliability and validity criteria; any data collected prior to meeting these criteria were discarded. At every testing site, either a member of the research team or a specially trained head assessor was present to oversee the testing and ensure fidelity of administration.

After each pilot testing window (i.e., fall 2009, spring 2010, fall 2010, and spring 2011), we conducted item analyses to identify poorly-performing items, which were then either revised or eliminated. In our task and item analyses, we examined both internal consistency and item discrimination. To assess item discrimination, we computed the discrimination index (D), using the two subgroups that comprised the highest and lowest 27 percent of the test sample in terms of total scores as recommended by Kelley (1939). D was the proportion of participants in the upper group that were correct less the proportion of participants in the lower group that were correct. We also computed the point-biserial correlation (i.e., item-total correlation coefficient). We flagged items as candidates for revision or removal that had both a point-biserial correlation of less than .30 (Englehart, 1965) and had a D of less than .30 (Reynolds, Livingston, & Willson, 2006). To assess internal consistency, we used Cronbach's alpha to measure how well each set of items measured a single unidimensional construct, with measures above .70 considered acceptable (Reynolds et al., 2006).

Field Testing

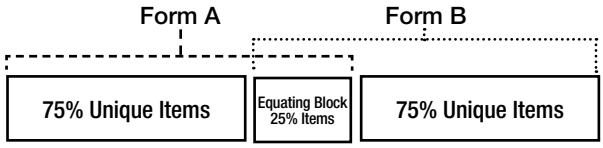
Construction of Parallel Forms

After analyzing all data from the 2009–2010 and 2010–2011 pilots, we assembled two tentative parallel forms of PALS español K. To assign items to the parallel forms, we rank ordered all items in the item pool for each individual subtask from easiest to most difficult based on the results of psychometric item analyses from the initial item pilot testing. We then used this rank-ordered list of items to assign items with approximately equal difficulty levels to each parallel form. (Because of the nature of the tasks, this process did not include the Alphabet and Digraph Recognition task or the Letter Sounds task).

Equating block. From the rank-ordered list of items covering the same subtask, we first chose items for the equating block that would be shared by the two parallel forms. These equating block items covered the full range of item difficulty values (i.e., easy to difficult).

Item assignment to two parallel forms. Once the equating block items were set aside, the remaining items in the same subtask were again rank ordered based on the item difficulty index values. We then used this list of rank-ordered items to assign items of

equal difficulty to the two forms. This procedure produced two parallel forms with approximately equal difficulty levels. The two parallel forms (Form A and Form B), with their shared common items (the equating block) and their respective unique items, are schematically represented below.



Field Test Data Collection

Our own trained assessors administered approximately 25% of the assessments in both years of field testing. The remaining 75% were administered by classroom teachers who had completed an interactive online training and certification module. In fall 2011, the tentative Form A of PALS español K was administered to 834 Spanish-speaking kindergartners in 32 schools in Virginia, Minnesota, California, Missouri, and Wisconsin. After revising or eliminating poorly performing items, we administered Form A again in spring 2012. Participants were 425 kindergartners in 15 schools in Virginia, Minnesota, Missouri, and Washington D.C. After the spring administration, we identified the best performing items for each task and eliminated the remainder to create a final version of Form A of PALS español K.

The tentative Form B of PALS español K was administered in fall 2012 to 562 Spanish-speaking kindergartners in 18 schools in Virginia, Wisconsin, Missouri, and Washington D.C. We then revised or eliminated poorly performing items

Table 2 Field Test Sample Demographics

Demographic Category		Fall 2011 *n* = 834	Spring 2012 *n* = 425	Fall 2012 *n* = 564	Spring 2013 *n* = 502
Gender	Male	.50	.50	.48	.45
	Female	.50	.50	.52	.55
Age	Mean Age in Months	66.8	73.6	67.9	74.5
FRPL	0% – 25%	.03	.09	.01	.01
	26% – 50%	.15	.13	.13	.17
	51% – 75%	.51	.58	.79	.63
	76% – 100%	.31	.20	.07	.19
Instructional Program	ESL	.46	.51	.31	.27
	Transitional Bilingual	.12	.00	.08	.12
	Dual Language	.19	.25	.43	.44
	Spanish Immersion	.23	.24	.18	.17

Note. Numbers other than age represent percentages. FRPL = Free or Reduced Price Lunch (used as proxy for SES).

before administering a revised version of tentative Form B to 502 kindergartners in spring 2013. The students in the spring 2013 field test were from 19 schools in Virginia, Illinois, Wisconsin, Missouri, and Washington D.C. After the spring administration, we once again identified the best performing items and eliminated the remainder to create a final version of Form B of PALS español K. Table 2 reports demographic information for the samples in each field test assessment period (i.e., fall 2011, spring 2012, fall 2012, and spring 2013). Note that the majority of the field tests included students participating in English as a Second Language (ESL), Dual Language, and Spanish Immersion programs; fewer students were enrolled in Transitional Bilingual programs.

The analyses we used after each field testing window (i.e., fall 2011, spring 2012, fall 2012, and spring 2013) to identify poorly performing items were the same analyses used for this purpose during pilot testing. We examined internal consistency using Cronbach's alpha, with measures above .70 considered acceptable (Reynolds et al., 2006), and we examined item discrimination using the discrimination index (D) and the point-biserial correlation (i.e., item-total correlation coefficient). As in the pilot phase, we flagged items as candidates for revision or removal that had both a point-biserial correlation of less than +.30 (Englehart, 1965) and had a D of less than .30 (Reynolds et al., 2006).

Advisory Board

Throughout this iterative process, we were guided by input from an Advisory Board, which provided recommendations on our task conceptualizations and gave advice on item development. Dr. Igone Arteagoitia, a researcher at the Center for Applied Linguistics in Washington, D.C., had oversight for the advisory board. Dr. Arteagoitia is a native Spanish speaker from Spain who holds a Ph.D. in Applied Spanish Linguistics.

The PALS español advisory board consisted of the following experts:

- **Dr. Vivian Correa**, a native Spanish speaker from Puerto Rico, whose research focuses on ELL special populations. Dr. Correa is a Professor in the Department of Special Education and Child Development at the University of North Carolina at Charlotte.
- **Dr. Verónica Galván Carlan**, a native Spanish speaker from Mexico, formerly an Assistant Professor in Early Childhood Education at the University of Texas at Brownsville and Texas Southmost College and currently an independent consultant.
- **Dr. Valerie Malabonga**, a developmental psychologist and consultant to the Center for Applied Linguistics in Washington, D.C., whose specialty is assessment, including literacy assessment, of bilingual children.
- **Dr. Lori Helman**, a specialist in second-language literacy development, who is an Associate Professor in the College of Education and Human Development at the University of Minnesota.
- **Ms. C. J. Grace**, formerly Bilingual Literacy Coordinator for St. Vrain Valley Schools in St. Vrain, Colorado and currently English Language Acquisition Coordinator for Denver Public Schools.

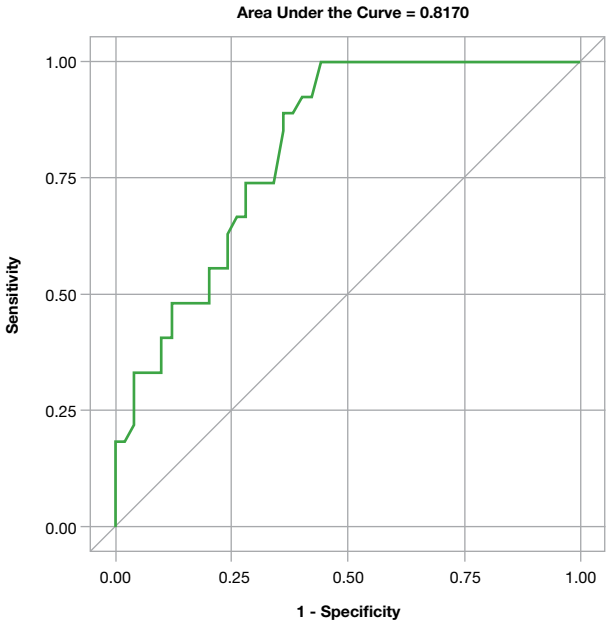
Section III

Establishing Summed Score Criteria and Benchmarks

Benchmarks for PALS español K were developed using data gathered from 2,325 administrations of the two forms of the assessment during field testing in 2011–2012 and 2012–2013. To derive the task benchmarks, which consist of raw scores, we began by examining task means and standard deviations for scores from all students above the bottom quartile. We established initial benchmarks by subtracting one standard deviation from the mean score for students in this group and then making adjustments based on modal data for each task. Finally, we evaluated the benchmarks subjectively to ensure that they reflect achievement milestones that are realistic and consistent with Spanish literacy acquisition theory and instructional practice.

The PALS español total score, called the Summed Score, is the sum of the benchmark scores for the core tasks. To test the accuracy of the PALS español K Summed Score in identifying students at risk for developing reading difficulties, we conducted

Figure 1 Area Under the Curve (AUC)



Spring 2012 PALS español K predicting risk status using spring 2013 Logramos Reading Total Score (n = 77).

Table 3 ROC Curve Analysis Studies

Assessments	Type	*n*	AUC	Discrimination[a]
PALS esp K → L-WA	Concurrent	84	.89	Excellent
PALS esp K → L-WA	Predictive[b]	98	.81	Excellent
PALS esp K → L-RT	Predictive[b]	77	.82	Excellent
PALS esp K → L-RT	Predictive[c]	73	.79	Acceptable
PALS esp K → MAP	Predictive[d]	76	.83	Excellent
PALS esp K → MAP	Concurrent	59	.81	Excellent
PALS esp K → MAP	Concurrent	77	.82	Excellent
PALS esp K → PALS K Eng	Concurrent	442	.89	Excellent

Note. L-WA = Logramos Word Analysis. L-RT = Logramos Reading Total. MAP = Measures of Academic Progress for Primary Grades. PALS K Eng = PALS K in English. The definition of risk using the outcome variable is those students at or below the 20% national percentile rank on the MAP or Logramos. Based on Hosmer & Lemeshow (1989). [b]1-year interval across grades. [c]2-year interval. [d]1-year interval within grade.

a series of receiver-operating characteristic (ROC) curve analyses. ROC curve analysis is a tool for evaluating how well an assessment classifies subjects into one of two categories, in this case being at risk or not being at risk for future reading difficulties. The Area Under the Curve (AUC) statistic of a ROC curve analysis is an indication of overall diagnostic accuracy (AUC values of 1.00 indicate perfect classification accuracy; values of .50 indicate accuracy no better than chance). Based on informal guidelines suggested by Hosmer and Lemeshow (1989), PALS español K has excellent discriminating capabilities. Table 3 reports AUC statistics from studies using PALS español K and three external indicators of risk: Logramos (2006), Measures of Academic Progress (MAP) for Primary Grades (2012), and PALS K in English. Note that the AUC values range from .79 to .89 for Spanish (Logramos) and from .81 to .89 for English (MAP and PALS K English) reading achievement tests. Figure 1 shows the AUC for spring 2012 PALS español predicting spring 2013 Logramos Total Reading.

Section IV
Technical Adequacy

Reliability

Reliability refers to the degree to which an assessment produces consistent results. We assessed three types of reliability for PALS español K: internal consistency, test-retest reliability, and inter-rater reliability.

Internal Consistency

Internal consistency measures the extent to which multiple items designed to measure the same construct produce similar scores. We used Cronbach's alpha to measure the internal consistency of each task on PALS español K. All tasks were found to have good internal consistency.

Rhyme Awareness and Beginning Sound Awareness. In both the rhyme and beginning sounds tasks, students are asked to match pictures, based on their phonological attributes. Analyses of internal consistency of these tasks yielded Cronbach's alphas of .84 and .81 for Rhyme Awareness and .83 and .79 for Beginning Sound Awareness in Forms A and B, respectively.

Alphabet and Digraph Recognition. The single best predictor of early reading achievement in Spanish is accurate, rapid naming of the letters of the alphabet (Bravo et al., 2006). The PALS español K alphabet recognition task requires students to point to and name 29 letters and digraphs. Analyses of internal consistency of the Alphabet and Digraph Recognition task yielded Cronbach's alphas of .98 for Form A and .97 for Form B.

Letter Sounds. In addition to naming the letters of the alphabet, young readers must develop knowledge of letter sounds and learn to apply that knowledge. The ability to produce the sounds represented by individual letters in isolation requires explicit awareness of individual phonemes, a requisite skill for reading development (Mathes et al., 2007). In the PALS español K Letter Sounds task, children are given a set of upper-case letters and asked to touch each letter and say the sound it represents. Note that children are given credit for either a single phoneme response (e.g., /*d*/ for the letter D) or a syllabic response (e.g., /*da*/, /*de*/, /*di*/, /*do*/, or /*du*/ for the letter d). Analyses of internal consistency yielded Cronbach's alphas of .87 for Form A and .91 for Form B of the Letter Sound Awareness task.

Concept of Word. Concept of word refers to the emergent reader's ability to match spoken words to written words as he/she reads (Clay, 1977; Henderson & Beers, 1980; Morris, 1981; Morris, Bloodgood, Lomax, & Perney, 2003; Roberts, 1992). Concept of word in text is a skill that precedes the ability to fully segment all the phonemes within words (Flanigan, 2007; Morris, 1993). Children with a solid concept of word will recognize words they did not know prior to reading a memorized or familiar text, even when these words are presented out of context. To administer the PALS español K Concept of Word tasks, the teacher assists the child in memorizing a nursery rhyme. Next, the teacher models reading the rhyme and then reads the rhyme chorally with the child. During both readings, the teacher points to each word as it is read. Finally, the child is asked to read the rhyme alone while pointing to each word. After the reading, the child is asked to point to targeted words within the text. The final task requires the child to read a list of ten words taken from the text of the rhyme. Analyses of internal consistency based on data from field testing the final forms of the Concept of Word task yielded Cronbach's alphas of .93 for Form A and .93 for Form B.

Spelling. Research on how children learn to read and spell words in an alphabetic orthography has

Table 4 Task Reliability (Cronbach's Alpha) by Instructional Program Type

Testing Window	ESL	Bilingual	Dual Language	Spanish Immersion
Form A				
Fall 2011	.95 (263)	.98 (95)	.97 (113)	.96 (169)
Spring 2012	.95 (210)	N/A (0)	.96 (103)	.96 (101)
Form B				
Fall 2012	.96 (178)	.96 (44)	.97 (242)	.96 (100)
Spring 2013	.98 (13)	.97 (59)	.96 (220)	.97(87)

Note. Subsample *n* in parentheses.

consistently revealed that orthographic features are internalized for reading and writing in a systematic, developmental progression (Bear, Templeton, Helman, & Baren, 2003; Defior, Jiménez-Fernández, & Serrano, 2005/2006; Diuk, Borzone, Sánchez Abchi, & Ferroni, 2009; Ferreiro, 1991; Hachén, 2002; Henderson & Beers, 1980; Jiménez et al., 2008; Sánchez-Abchi, Diuk, Borzone, & Ferroni, 2009). Invented spellings provide a diagnostic window into students' understanding of alphabetic orthography and can help teachers determine when to teach particular phonics or spelling features (Henderson, 1990; Invernizzi, Abouzeid, & Gill, 1994). In the PALS español K Spelling task, words were selected based on frequency of occurrence and linguistic attributes, as well as their ability to elicit responses to particular speech sounds and syllable patterns that represent developmental spelling stages in early Spanish literacy acquisition (Defior et al., 2009; Ferreiro, 1991; Ferreiro & Teberosky, 1982; Hachén, 2002). Analyses of internal consistency based on data from field testing the final forms of this task yielded Cronbach's alphas of .88 for Form A and .83 for Form B.

Subtask Reliability by Instructional Program

We further explored the reliability of PALS español K subtasks by using Cronbach's alpha to measure the internal consistency of the tasks as a whole across instructional program types (i.e., ESL, Transitional Bilingual, Dual Language, and Spanish Immersion). Coefficients were consistently high across all program types for both Form A and Form B, suggesting that PALS español K is a reliable measure of early literacy skills for students being taught in Spanish, in English, or in both languages simultaneously (see Table 4).

Test-Retest Reliability

The test-retest reliability coefficient provides an estimate of the stability of measurement outcome across two occasions with a specified time interval. To obtain test-retest reliability data for PALS español K, a sub-sample of students was selected for a second administration of PALS español K after an interval of between one and two weeks. Pearson product moment correlation analysis based on the PALS español K Summed Score yielded a correlation coefficient of .96 (n = 56).

Inter-Rater Reliability

Inter-rater reliability coefficients indicate how consistently different individuals score a particular task in the same manner. To determine the inter-rater reliability of PALS español K, two different raters scored the same task (i.e., one person administered and scored the task while a second person observed and scored the task simultaneously but independently). The scores of the two raters were then compared and measured using an intraclass correlation coefficient (ICC; Shrout & Fleiss, 1979). Intraclass correlation coefficients for PALS español tasks ranged from .73 to .998 across two separate inter-rater reliability studies (see Table 5).

Validity

Validity refers to how well an assessment measures what it is intended to measure. We assessed three types of validity for PALS español K, using data gathered during field testing in 2011–2012 and 2012–2013: content validity, construct validity, and criterion-related validity.

Content Validity

Content validity refers to how well the items and tasks included on an assessment provide a relevant and representative sample of the content being assessed (Gronlund, 1985). To ensure content validity of PALS español K, tasks and items were designed to assess the specific early literacy skills that research has shown to contribute to reading development in Spanish. Specifically, PALS español K measures children's progress toward developing phonological awareness, alphabet knowledge, and orthographic knowledge.

PALS español K phonological awareness tasks (i.e., Rhyme Awareness and Beginning Sound Awareness) measure students' ability to attend to the sounds that make up language at the word and phoneme levels. Items on both tasks consist of high-frequency words with concrete meanings that can be represented pictorially. Care was taken to ensure that the pictured items are representative of the content being assessed and are age and culture appropriate.

PALS español K alphabet knowledge tasks (i.e., Alphabet and Digraph Recognition and Letter Sounds) measure children's knowledge of the form and sounds of the letters and digraphs used in Spanish writing. To ensure that the alphabet knowledge tasks would have ample content validity, we included the complete set of letters and digraphs that are commonly taught in early literacy instruction. For example, although the digraphs ch and ll are no longer considered part of the official Spanish alphabet, we included them in the Alphabet and Digraph Recognition task to reflect current instructional practice. In the Letter Sounds task, we included both letters and digraphs, omitting only letters that have sounds that are difficult to produce individually (e.g., Q) and letters with ambiguous pronunciations (e.g., X).

To assess concept of word in text, we included subtasks at three levels of sophistication to reflect the steps children typically go through as they gradually master this skill. Children are first asked to touch each word in the text of a memorized rhyme as they recite the rhyme. Next, they are asked to identify specific words pointed out in the context of the text. Finally, in the most challenging subtask, they are

Table 5 Inter-Rater Reliability

PALS español K Task	ICC
Group Rhyme Awareness	.997 – .998 (184)
Individual Rhyme Awareness	.993 (91)
Group Beginning Sound Awareness	.979 – .989 (183)
Individual Beginning Sound Awareness	.988 (90)
Alphabet/Digraph Recognition	.856 – .991 (108)
Letter Sounds	.730 – .958 (106)
Spelling	.908 – .991 (213)
Concept of Word	.850 – .980 (92)

Note. ICC = Intraclass Correlation Coefficient. Sample *n* in parentheses.

asked to identify words from the text presented out of order in isolation.

Finally, the Spelling task on PALS español K is designed to assess children's progress toward learning to represent phonemes in one- and two-syllable words made up primarily of open syllables. The open syllable is the predominant syllable structure in Spanish (Delattre & Olsen, 1969) and forms the basis for initial reading and writing instruction in Spanish.

Additional information on the content validity of PALS español K tasks can be found in this section of the technical reference under Internal Consistency.

Criterion-related Validity

Analyses of criterion-related validity determine whether scores on an assessment are related to scores on one or more outcome criteria (AERA, APA, & NCME, 1999). There are two types of criterion-related validity: predictive validity, in which scores on an assessment are used to predict performance on another assessment administered in the future; and concurrent validity, in which scores on an assessment are compared to scores on another assessment administered at approximately the same time. We assessed both types of criterion-related validity for PALS español K.

Predictive Validity. To assess the predictive validity of PALS español K, we first examined the relationship between fall PALS español K scores and scores on another Spanish assessment (Logramos, 2006), administered the following spring, for a one-year, within grade level interval. Logramos is a norm-referenced assessment for Grades K–12 that measures skills in reading, language, and mathematics. We used the Logramos Word Analysis section for kindergarten, which measures alphabet recognition, beginning sound awareness, letter-sound relationships, and word and syllable knowledge, the skills most similar to the skills assessed by PALS español K. Fall PALS español K Summed Scores and fall PALS español K subtask scores were all significantly correlated ($p < .05$) with Logramos Word Analysis scores obtained at the end of the kindergarten year (see Table 6).

We also examined the relationship between fall PALS español K and Logramos, administered in fall and spring of first grade. For first-grade Logramos, we used the Total Reading Score, which is derived from the Vocabulary and Reading Comprehension tasks. All PALS español K subtasks were significantly correlated ($p < .05$) with the Logramos Total Reading Score at both one- and two-year across-grade intervals (see Table 6).

Table 6 Predictive Validity Studies

PALS español K Tasks/Scores	Logramos			MAP[a]
	L-WA[a]	L-RT[b]	L-RT[c]	
Summed Score	.66 (99)	.59 (77)	.46 (73)	.70 (76)
Rhyme Awareness	.42 (99)	.29 (77)	.27 (73)	.54 (76)
Beginning Sounds	.52 (98)	.50 (77)	.51 (73)	.54 (76)
Alphabet/Digraph Recognition	.69 (101)	.51 (77)	.37 (73)	.57 (76)
Letter Sounds	.42 (100)	.50 (77)	.34 (73)	.54 (76)
COW	.66 (103)	.43 (77)	.44 (73)	.49 (76)
Spelling	.75 (102)	.53 (77)	.26 (73)	.54 (76)

Note. L-WA = Logramos Word Analysis. L-RT = Logramos Reading Total. MAP = Measures of Academic Progress for Primary Grades. COW = Concept of Word. Sample *n* in parentheses.
[a]1-year interval within grade. [b]1-year interval across grades. [c]2-year interval.
All correlations are statistically significant ($p < .05$).

In both the one-year within-grade and the one- and two-year across-grade interval studies, the variables most highly correlated with Logramos were the overall Summed Score, beginning sounds, and the concept of word in text task (see Table 6).

Finally, because early literacy skills are known to transfer between languages (Durgunoglu, 1993), we explored the relationship between PALS español K and MAP, a literacy assessment administered in English. The fall PALS español K Summed Score and all task scores were found to be significantly correlated ($p < .05$) with MAP scores administered in spring of kindergarten. As can be seen in the right-hand column of Table 6, the one-year within-grade interval correlations between PALS español K and MAP were all in the moderate to high range.

Concurrent Validity. To assess the concurrent validity of PALS español K, we examined the relationship between scores on PALS español K and scores on three other assessments administered at approximately the same time: Logramos (Word Analysis), ¡Prepárate a leer! (Whitehurst & Lonigan, 2009), and MAP. Like Logramos, ¡Prepárate a leer! is administered in Spanish. As can be seen in Table 7, the PALS español K Summed Score and all subtask scores were significantly correlated ($p < .05$) with all three assessments. The highest correlations were found between the overall Summed Score, concept of word, and spelling, regardless of whether the outcome assessment was administered in Spanish (Logramos) or English (MAP).

We also tested the diagnostic accuracy (concurrent and predictive) of PALS español K tasks, using teacher evaluations of students' reading skill as the external measure. Prior to the two spring field testing windows (i.e., spring 2012 and spring 2013), we asked a subset of teachers to identify (1) the children who were in the top 25% of the class in terms of Spanish reading proficiency, and (2) the children who were in the bottom 25% of the class in terms of Spanish reading proficiency. We then used PALS español K data collected in the same testing window and in the previous testing window to evaluate the classification accuracy of PALS español K, based on teacher judgment. Using logistic regression to predict whether the child was in the top or bottom of the class, the outcome was regressed on the PALS español K Summed Score. Classification accuracy for concurrent predictions ranged from 86% to 92% ($n = 130$), and classification accuracy for future predictions (i.e., from fall to spring) ranged from 84% to 87% ($n = 110$). The results from these analyses suggest that PALS español K aligns with teacher judgment of students' relative standing in reading achievement.

Table 7 Concurrent Validity Studies

PALS español K Tasks/Scores	PL — Fall 2012	L-WA — Spring 2013	MAP — Spring 2013
Summed Score	.43 (125)	.82 (84)	.56 (77)
Rhyme Awareness	.35 (125)	.41 (86)	.47 (77)
Beginning Sounds	.42 (125)	.37 (88)	.51 (77)
Alphabet/Digraph Recognition	.26 (125)	.78 (87)	.43 (77)
Letter Sounds	.37 (125)	.67 (88)	.39 (77)
COW	.16 (125)	.76 (87)	.56 (77)
Spelling	.45 (125)	.66 (85)	.54 (77)

Note. PL = Prepárate a leer. L-WA = Logramos Word Analysis. MAP = Measures of Academic Progress for Primary Grades (Reading Test). COW = Concept of Word. Sample *n* in parentheses.
All correlations are statistically significant ($p < .05$).

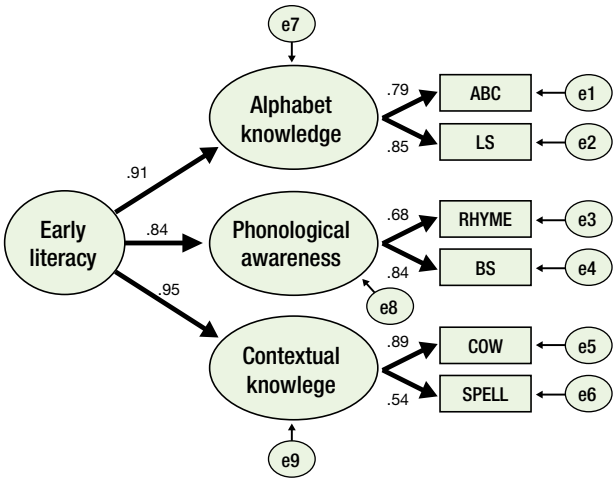
Construct Validity

Construct validity refers to the degree to which the underlying traits of an assessment can be identified and the extent to which these traits reflect the theoretical model on which the assessment was based (Gronlund, 1985).

PALS español K is based on a theoretical model that defines reading and writing as the application of both sound and print knowledge. Children must develop the ability to recognize and manipulate the sounds of the language, as well as an understanding about how print works as prerequisites for learning to read and write. PALS español K includes tasks that assess children's knowledge of sound and print, as well as composite tasks that represent a combination of the two. We tested this theoretical model using factor analysis and also by examining the intercorrelations among PALS español subtasks.

Factor Analysis. We investigated the latent factor structure of PALS español-K using data from the 2011–2012 field test (n = 590). Three theoretically-guided factor structures were measured and tested with one half of our sample using confirmatory factor analytic (CFA) methods. We then replicated the best fitting model with the second half of our sample (see Figure 1) and used multigroup CFA (MGCFA) to test for the model's configural and metric invariance. The results of the MGCFA showed a good fit with the data. All of the first-order factor loadings demonstrated that the six tasks are all good measures of their corresponding factors. Our findings support that one general factor, Early Literacy, gives effects to the other first-order factors. The MGCFA also revealed that PALS español-K exhibits measurement invariance when comparing populations of boys and girls (Huang, Ford, Invernizzi, & Fan, 2013). These results support the educational utility of PALS español-K as a tool for assessing important reading constructs and informing early interventions across gender groups with Spanish-speaking students.

Figure 2 Standardized factor loadings for best-fitting model (n = 294)



Note. ABC = alphabet/digraph recognition. LS = letter sound knowledge. RHYME = rhyme awareness. BS = beginning sound awareness. COW = concept of word. SPELL = spelling.

Table 8 Intercorrelations of PALS español K Tasks and Summed Score (Form A)

	1	2	3	4	5	6	7
1. Rhyme Awareness	1	0.50	0.45	0.40	0.53	0.48	0.60
2. Beginning Sound Awareness	0.50	1	0.44	0.45	0.55	0.51	0.62
3. Alphabet/Digraph Recognition	0.45	0.44	1	0.57	0.68	0.71	0.90
4. Letter Sounds	0.40	0.45	0.57	1	0.67	0.61	0.77
5. Spelling	0.53	0.55	0.68	0.67	1	0.80	0.89
6. COW Word List	0.48	0.51	0.71	0.61	0.80	1	0.86
7. Summed Score	0.60	0.62	0.90	0.77	0.89	0.86	1

Note. Correlations are based on analysis of complete cases. COW = Concept of Word. n = 414 for all tasks.
All correlations are statistically significant ($p < .01$).

Table 9 Intercorrelations of PALS español K Tasks and Summed Score (Form B)

	1	2	3	4	5	6	7
1. Rhyme Awareness	1	0.41	0.31	0.35	0.45	0.48	0.53
2. Beginning Sound Awareness	0.41	1	0.32	0.45	0.56	0.56	0.60
3. Alphabet Recognition	0.31	0.32	1	0.56	0.44	0.58	0.83
4. Letter Sounds	0.35	0.45	0.56	1	0.65	0.69	0.82
5. Spelling	0.45	0.56	0.44	0.65	1	0.79	0.81
6. COW Word List	0.48	0.56	0.58	0.69	0.79	1	0.87
7. Summed Score	0.53	0.60	0.83	0.82	0.81	0.87	1

Note. Correlations are based on analysis of complete cases. COW = Concept of Word. $n = 495$ for all tasks.
All correlations are statistically significant ($p < .01$).

Intercorrelations. We conducted Pearson product moment correlation analysis to examine the relationships between the kindergarten tasks and the overall Summed Score for Form A in spring 2012 and Form B in spring 2013. Correlations between all tasks were significant and in the moderate to high range ($r = .40–.80$ for Form A and .31–.79 for Form B, $p < .001$). The correlations that were consistently high in both Form A and Form B were between Letter Sounds and Spelling, between Letter Sounds and Concept of Word, and between Spelling and Concept of Word. Each of the kindergarten tasks was also found to be significantly correlated with the overall Summed Score ($r = .60–.90$ for Form A and .53–.87 for Form B, $p = < .001$). The highest correlations with the overall Summed Scores for both forms included alphabet recognition, letter sounds, spelling, and concept of word. Intercorrelations are reported in Tables 8 and 9.

Section V
Summary

The technical adequacy of PALS español K has been established through 2,325 administrations of the two forms of the assessment during field testing in 2011–2012 and 2012–2013. The reliability of PALS español subtasks has been demonstrated through the use of Cronbach's alpha. Reliability coefficients for individual tasks range from .79 to .98, demonstrating good internal consistency. Inter-rater reliabilities expressed as intraclass correlation coefficients ranged from .73 to .998, indicating that PALS español K can be scored consistently across individuals.

Analyses of field test data also support the content, construct, and criterion-related validity of PALS español K. Factor analysis and intercorrelations among subtasks have demonstrated the construct validity of PALS español. Correlation analyses have provided evidence of concurrent and predictive validity, showing that PALS español K scores are significantly correlated with scores on both Spanish and English literacy measures administered at approximately the same time and at one- and two-year intervals. ROC curve analysis and logistic regression analysis have demonstrated the diagnostic accuracy of PALS español in identifying children at risk for developing reading difficulties.

In summary, PALS español K is an assessment tool with good evidence of reliability and validity that can be used effectively to screen Spanish-speaking students in kindergarten for possible reading difficulties. PALS español K shows evidence of internal consistency, demonstrating that the items within each task produce similar results in measuring the same construct. PALS español K also shows evidence of good inter-rater and test-retest reliability, showing that the assessment can be administered and scored consistently by different users and that it is a stable measure across time. PALS español K also demonstrates good evidence of content, construct, and criterion-related validity, suggesting that PALS español K truly measures the underlying constructs associated with early literacy development in Spanish.

Section VI
References

American Educational Research Association (AERA), American Psychological Association (APA), & National Council on Measurement in Education (NCME). (1999). Standards for educational and psychological testing. Washington D.C.: American Psychological Association.

Bear, D. R., Templeton, S., Helman, L. A., & Baren, T. (2003). Orthographic development and learning to read in two different languages. In G. G. García (Ed.), *English learners: Reaching the highest level of English literacy* (pp. 71–95). Newark, DE: International Reading Association.

Branum-Martin, L., Mehta, P. D., Fletcher, J. M., Carlson, C. D., Ortiz, A. A., Carlo, M. S., & Franis, D. J. (2006). Bilingual phonological awareness: Multilevel construct validation among Spanish-speaking kindergarteners in transitional bilingual education classrooms. *Journal of Educational Psychology*, 98, 170–181.

Bravo, L. (1995). A four year follow-up study of low socioeconomic status, Latin American children with reading difficulties. *International Journal of Disability, Development and Education*, 42, 189–202.

Bravo, L., Villalón, M., & Orellana, E. (2006). Diferencias en la predictividad de la lectura entre primer año y cuarto año básicos, Psykhe, 15, 3–11.

Carrillo, M. (1994). Development of phonological awareness and reading acquisition. *Reading and Writing: An Interdisciplinary Journal, 6*, 279–298.

Clay, M. (1977). *Reading: The patterning of complex behavior*. Exeter, NH: Heinemann Educational Books.

Cuetos, F., & Suárez-Coalla, P. (2009). From grapheme to word in reading acquisition in Spanish. *Applied Psycholinguistics*, 30, 583–601.

Davies, R., Cuetos, F., & Glez-Seijas, R. M. (2007). Reading development and dyslexia in a transparent orthography: A survey of Spanish children. *Annals of Dyslexia, 57*, 179–198.

Defior, S., Jiménez- Fernández, G., & Serrano, F. (2005/2006). Spelling acquisition: A transveral study of Spanish children. *International Journal of Learning, 12*, 293–299.

Delattre, P., & Olsen, C. (1969). Syllabic features and phonic impression in English, German, French and Spanish. *Lingua, 22*, 160–175.

Diuk, B., Borzone, A. M., Sánchez Abchi, V., & Ferroni, M. (2009). La adquisición de conocimiento ortográfico en niños de 1er a 3er año de educación básica. *Psykhe, 18*, 61–71.

Durgunoglu, A. Y., Nagy, W. E., & Hancin-Bhatt, B. J. (1993). Cross-language transfer of phonological awareness. *Journal of Educational Psychology, 85*, 453–465.

Englehart, M. D. (1965). A comparison of several item discrimination indices. *Journal of Educational Measurement, 2*, 69–76.

Escribano, C. L. (2007). Evaluation of the double-deficit hypothesis subtype classification of readers in Spanish. *Journal of Learning Disabilities, 40*, 319–330.

Escribano, C. L., Elosúa, M. R., Gómez-Veiga, I., & García Madruga, J. A. (2013). A predictive study of reading comprehension in third-grade Spanish students. *Psicothema, 25*, 199–205.

Ferreiro, E. (1991). La construcción de la escritura en el niño. Lectura y vida: Revista latinoamericana de lectura, 12(3), 5–14.

Ferreiro, E., & Teberosky, A. (1982). Literacy before schooling (K. G. Castro, Trans.). Exeter, NH: Heinemann Educational Books.

Flanigan, K. (2007). A concept of word in text: A pivotal event in early reading acquisition. Journal of Literacy Research, 39, 37–70.

Gómez, L. F. (2008). El desarollo de la competencia lectora en los primeros grados de primaria. Revista Latinoamericana de Estudios Educativos (México), 38 (3–4), 95–126.

Goswami, U. (2010). Typical reading development and developmental dyslexia across languages. In D. Coch, G. Dawson, & K. W. Fischer (Eds.), Human behavior, learning and the developing brain: Atypical development (pp. 145–167). New York, NY: Guilford Press.

Gronlund, N. E. (1985). *Measurement and evaluation in teaching*. New York, NY: Macmillan.

Hachén, R. (2002). Conocimiento lingüístico y reflexión metalingüística: El rol de las conceptualizaciones en torno a la estructura de la sílaba en el proceso de alfabetización. *Lectura y vida: Revista latinoamericana de lectura, 23*(2), 6–17.

Henderson, E. (1990). *Teaching spelling* (2nd ed.). Boston: Houghton Mifflin.

Henderson, E., & Beers, J. (1980). *Developmental and cognitive aspects of learning to spell*. Newark, DE: International Reading Association.

Herrera, L., & Defior, S. (2005). Una aproximación al procesamiento fonológico de los niños prelectores: Conciencia fonológica, memoria verbal a corto plazo y denominación. *Psykhe, 14*, 81–95.

Hosmer, H. W., Jr., & Lemeshow, S. (1989). *Applied Logistic Regression*. New York, NY: John Wiley & Sons.

Huang, F., Ford, K., Invernizzi, M., & Fan, X. (2013, April). Investigating the factor structure of the Phonological Awareness Literacy Screening in Spanish for Kindergartners (PALS español K) using hierarchical confirmatory factor analysis. Poster session presented at the meeting of the American Educational Researchers Association, San Francisco, CA.

Invernizzi, M., Abouzeid, M., & Gill, J. T. (1994). Using students' invented spellings as a guide for spelling instruction that emphasizes word study. *The Elementary School Journal, 95*, 155–167.

Jiménez, J. E., & Ortiz, M. (2000). *Conciencia fonológica y aprendizaje de la lectura: Teoría, evaluación e intervención.* Madrid: Editorial Síntesis.

Jiménez, J. E., O'Shanahan, I., Tabraue, M., Artiles, C., Muñetón, M., Guzmán, R. . . ., & Rojas, E. (2008). Evolución de la escritura de palabras de ortografía arbitraria en lengua española. *Psicothema, 20*, 786–794.

Kelley, T. L. (1939). The selection of upper and lower groups for the validation of test items. *Journal of Educational Psychology, 30*, 17–34.

Logramos (2nd Ed). (2006). Rolling Meadows, IL: Riverside.

Manis, F. R., Lindsey, K. A., & Bailey, C. E. (2004). Development of reading in grades K–2 in Spanish-speaking English-language learners. *Learning Disabilities Research & Practice, 19*, 214–224.

Manrique, A. M. B., & Signorini, A. (1998). Emergent writing forms in Spanish. *Reading and Writing: An Interdisciplinary Journal, 10*, 499–517.

Mathes, P. G., Pollard-Durodola, S., Cárdenas-Hagan, E., Linan-Thompson, S., & Vaughn, S. (2007). Teaching struggling readers who are native Spanish speakers: What do we know? *Language, Speech, and Hearing Services in Schools, 38*, 260–271.

Measures of Academic Progress (MAP) for Primary Grades. (2012). Portland, OR: Northwest Evaluation Association.

Morris, D. (1981). Concept of word: A developmental phenomenon in the beginning reading and writing process. *Language Arts, 58*, 659–668.

Morris, D. (1993). The relationship between children's concept of word in text and phoneme awareness in learning to read: A longitudinal study. *Research in the Teaching of English, 27*, 133–154.

Morris, D., Bloodgood, J., Lomax, C., & Perney, J. (2003). Developmental steps in learning to read: A longitudinal study in kindergarten and first grade. *Reading Research Quarterly, 38*, 302–328.

Pollard-Durodola, S., & Simmons, D. C. (2009). The role of explicit instruction and instructional design in promoting phonemic awareness development and transfer from Spanish to English. *Reading and Writing Quarterly, 25*, 139–161.

Reynolds, C. R., Livingston, R. B., & Willson, V. (2006). *Measurement and assessment in education*. Boston, MA: Allyn and Bacon.

Roberts, E. (1992). The evolution of the young child's concept of word as a unit of spoken and written language. *Reading Research Quarterly, 27*, 124–138.

Sánchez Abchi, V., Diuk, B., Borzone, A. M., & Ferroni, M. (2009). El desarrollo de la escritura de palabras en español: Interacción entre el conocimiento fonológico y ortográfico. *Interdisciplinaria, 26*, 95–119.

Shrout, P., & Fleiss, J. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin, 86*, 420–428.

Whitehurst, G. J., & Lonigan, C.J. (2009). *¡Prepárate a leer!: Herramienta de observacion folleto para ninos, revisada.* New York: Pearson Early Learning.

Youden, W. J. (1950). An index for rating diagnostic tests. *Cancer, 3*, 32–35.