pals™ español

Phonological Awareness Literacy Screening

Grades 1–3

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

1–3 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 1–3 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 1–3. 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 1–3 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

Introduction

Purpose and Uses

The primary purpose of PALS español 1–3 is to identify students who are performing below grade-level expectations in fundamental literacy skills and therefore may be in need of additional literacy instruction beyond what is provided to typically developing readers. PALS español 1–3 also serves as a diagnostic tool that allows teachers to target instruction to each student's individual literacy needs. PALS español 1–3 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 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

PALS español 1–3 measures children's progress toward developing important Spanish literacy fundamentals in the areas of: (a) alphabet and digraph knowledge, (b) knowledge of letter sounds, (c) phoneme-grapheme correspondences, (d) phonemic awareness, (e) concept of word in text, (f) orthographic knowledge, (g) word recognition in isolation, and (h) word recognition in context, including accuracy, fluency, and comprehension. As shown in Table 1, PALS español tasks are organized into four levels. All students are administered the Entry Level and Level A tasks. Only students who do not meet the benchmark on the Entry Level tasks are administered Level B tasks. Those who do not meet the benchmarks on the Level B tasks are also administered Level C.

Level B and Level C tasks are designed to allow teachers to determine whether students who are not preforming according to expectations need additional help with foundational literacy skills in the domains of phonemic awareness, alphabet knowledge, and orthographic knowledge. *Phonemic awareness* refers to the ability to manipulate oral language at the phoneme level (e.g., blending and segmenting individual sounds). *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*). Research has clearly demonstrated the importance of skills in each of these areas in predicting later reading achievement in Spanish among both monolingual and bilingual Spanish-speaking children (Branum-Martin et al., 2006; Bravo, 1995; Bravo, Villalón, & Orellana, 2006; Davies, Cuetos, & Glez-Seijas, 2007; Escribano, 2007; Escribano, Elosúa, Gómez-Veiga, & García-Madruga, 2013; Gómez, 2008; Goswami, 2010; Herrera & Defior, 2005; Jiménez & Ortiz, 2000; Manis, Lindsey, & Bailey, 2004).

In a study that explored the importance of phonemic 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

Table 1 PALS español 1–3 Tasks

Tasks		Domains			
		Orthographic Knowledge	Word Recognition in Context	Alphabet Knowledge	Phonemic Awareness
Entry Level	Word Recognition in Isolation	✓		✓	✓
	Spelling	✓		✓	✓
Level A	Oral Reading in Context	✓	✓	✓	✓
Level B	Alphabet/Digraph Recognition			✓	
	Letter Sounds	✓		✓	
	Concept of Word in Text	✓		✓	✓
Level C	Blending				✓
	Sound-to-Letter			✓	✓

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. In a study with 120 kindergarten and first grade children in Murcia, Spain, Carrillo (1994) examined the relationship between different levels of phonological awareness skills (i.e., less to more challenging) and learning to read in Spanish. She found that the ability to segment speech at the phoneme level was associated with the development of decoding skills and clearly differentiated good readers from poor to average readers. 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. These studies support the theoretical rationale for the tasks on PALS español that are heavily reliant on phonological and orthographic knowledge.

PALS español 1–3 Administration and Scoring

PALS español 1–3 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 Spelling task, which can be administered to an entire class at once. Students receive a score on each PALS español 1–3 task, as well as overall scores, called Summed Scores, on the Entry Level tasks and the Level B tasks. The Summed Scores are calculated by adding together students' scores on the tasks that make up those levels (i.e., Spelling and Word Recognition in Isolation for the Entry Level Summed Score; Alphabet and Digraph Recognition,

Letter Sounds, and Concept of Word for the Level B Summed Score).

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 and can design instruction accordingly. Benchmarks are also provided for the PALS español 1–3 Entry Level Summed Score and Level B Summed Score for fall and spring. The Entry Level Summed Score benchmark is used to identify students who may be at risk for developing reading difficulties. The Level B Summed Score benchmark is used to determine which students should be administered the additional diagnostic tasks in Level C that focus more deeply on phonemic awareness.

Detailed procedures for administering and scoring PALS español 1–3 can be found in the Administration and Scoring Guide. A description of how PALS español 1–3 benchmarks were established can be found in Section III of this Technical Reference.

Section II

Item Development and Field Testing

Preliminary Test Development

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. Below is a description of each of the tasks that make up PALS español 1–3 as well as a description of how we selected the items for each task.

Spelling. The PALS español Spelling task is a developmental spelling inventory designed to assess students' understanding of nine distinct Spanish spelling features (i.e., open syllables, closed syllables, blends, nasals, diphthongs, inconsistent consonants, rule-based consonants, silent H, and affixes/roots) that represent a continuum from easiest to most difficult. All words on the Spelling task were selected based on frequency of occurrence and linguistic attributes, as well as their ability to elicit responses to particular speech sounds and word patterns that represent developmental spelling stages in Spanish literacy acquisition (Defior et al., 2009; Ferreiro, 1991; Ferreiro & Teberosky, 1982; Hachén, 2002).

Words are grouped into three sets, based on the nature of the spelling features and their level of difficulty. Set 1 words are the easiest to spell because they can be spelled by relying simply on knowledge of letter-sound correspondences (e.g., mesa, gato). Set 2 words are more difficult because their spelling requires an understanding of word structures or patterns (e.g., lápices, guitarra). Set 3 words are the most difficult to spell because they require students to have knowledge of the words' derivations in order to be sure of their spelling (e.g., geografía, inmenso).

Word Recognition in Isolation. Words were chosen from sets of grade-level words harvested from Spanish-language texts and were assigned to lists based on the relative difficulty of the syllable and word patterns they represented, as well as their standard frequency indices. The standard frequency index (SFI) of a word represents the frequency with which that word might be expected to occur in Spanish. Word frequency has been shown to be associated with level of difficulty in both reading and spelling. We computed the SFI using the following formula: SFI = 10 * [log10(U) + 4], in which U represents the estimated number of occurances per one million words. We obtained the estimated occurances per million words from the *Corpus del español*, a 100 million-word corpus of Spanish words collected from print and oral sources (Davies, 2002). A word with an SFI of 90

would be expected to occur once in every 10 words; one with an SFI of 80 would be expected to occur once in every 100 words; one with an SFI of 70 would be expected to occur once in every 1000 words, etc. The average SFI indices for the PALS español word lists are as follows: Preprimer – 61, Primer – 60, 1st grade – 58, 2nd grade – 56, and 3rd grade – 54.

Oral Reading in Context. Listening to students read aloud from graded passages provides direct information for estimating reading levels, diagnosing strengths and weaknesses, and evaluating progress (Johnson, Kress, & Pikulski, 1987). PALS español 1–3 includes nine reading passages, ranging from Readiness level to fourth grade. The first-grade through fourth-grade passages are expository passages. Expository passages were used to avoid cultural bias in narrative expectations (Barr, 1984). In writing the passages, we referred to grade-level word lists created from elementary-level Spanish texts. In addition to using word frequency, we also used text features that are known to be associated with difficulty level. These features include the total number of words in the passage, the number of words per sentence and per page, and the number of sentences per page, as well as linguistic patterns, lexical density, and predictability (for the Preprimer and Primer passages). We used Lexile analyses (Metametrics: Spanish Lexile Analyzer) to verify the relative difficulty level of the passages and then empirically validated the lexiles by having students reading three successive levels to verify that oral reading accuracy and oral reading rate decreased with each increment in difficulty. Table 2 reports lexile measures for the Primer through fourth-grade passages.

Table 2 Lexile Measures of PALS español Oral Reading in Context Passages

Passage Level	Lexile Measures	
	Form A	Form B
Primer	300L	320L
1st Grade	480L	460L
2nd Grade	540L	540L
3rd Grade	660L	680L
4th Grade	790L	760L

Oral Reading in Context is assessed for accuracy, fluency, and comprehension. *Accuracy* on oral reading in context is measured as the percentage of words read correctly in a leveled passage. *Fluency* is assessed in two ways: (a) through a three-level rating scale adapted from the National Assessment of Educational Progress (NAEP) and (b) through measures of oral reading rate. *Comprehension* is assessed for each reading passage (Primer through 4th grade) by means of 10 text-dependent questions that represent a balance of factual, main idea, inference, and vocabulary questions. At the primer and 1st-grade levels, teachers read the questions and possible answers aloud, and the student is asked to provide answers on paper. At the 2nd- through 4th-grade levels, students are asked to read the questions and the answer choices silently, and then respond on paper.

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 Alphabet and Digraph Recognition task requires students to point to and name all 27 letters of the alphabet. We also included the digraphs *ch* and *ll* because they are essential to Spanish reading and are typically included in beginning reading instruction.

Letter Sounds. In the PALS español Letter Sounds task, children are given a set of upper-case letters and digraphs and asked to touch each one and say the sound it represents. 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). As in the Alphabet and Digraph Recognition task, we included the digraphs ch and ll because they are typically part of beginning reading instruction.

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 Concept of Word task, 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. The rhymes used in PALS español are traditional nursery rhymes approved by our advisory board of native Spanish speakers and literacy specialists.

Blending. PALS español includes two measures of phonological awareness at the phoneme level: (a) a phoneme blending task and (b) a phoneme segmentation task. Research (Branum-Martin et al., 2006; Bravo, 1995; Carrillo, 1994) has shown that the ability to manipulate language at the phoneme level is associated with successful reading in Spanish. The Blending task is a phonological processing task that requires students to use information from the sound structure of speech to retrieve words. When administering the task, the teacher vocalizes specific sounds and asks the student to put them together and identify a word. For example, the teacher might say "/d/ /o/ /s/," and the student would respond by blending the sounds together to make the word *dos*. The item sequence for this task follows a developmental progression from easy to more difficult; children are first asked to blend two-letter words, followed by three-letter words, and finally four-letter words within consonant blends.

Sound-to-Letter. The Sound-to-Letter task also measures students' phonological awareness at the phoneme level but also assesses students' ability to apply that phonemic awareness to the letters of the alphabet. For this task, students are required to segment words into their constituent phonemes and then identify the letter that corresponds to one of the phonemes (i.e., the beginning sound, the middle sounds, or the final sound) in the word. The Sound-to-Letter task follows a developmental progression from easy to more difficult and assesses the depth of students' understanding of the alphabetic principle. Children are first asked to segment the initial phoneme, then the final phoneme, and finally, the medial phoneme.

Initial Piloting

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 of Items

During the 2009–2010 and 2010–2011 academic years, we conducted pilot tests of all PALS español 1–3 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 Blending 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 students in Grades 1–3 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 1–3 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 for all tasks except Oral Reading in Context to identify poorly-performing items, which were then either revised or eliminated. (Reading passages were not reviewed in the same manner because the nature of the task precluded the removal of items from the stories.) 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).

To analyze the reading passages in the Oral Reading in Context task, we computed item difficulties (p) per word to ensure that there were no words that were problematic for readers at all levels (i.e., strong

readers as well as weak readers). We also conducted incremental level of difficulty analyses to ensure that each higher grade-level passage was marked by a statistically significant decline in words correct per minute (WCPM; combining the factors of accuracy and words per minute), demonstrating an increasing level of difficulty. Paired sample *t*-tests were conducted with a *Bonferroni* corrected alpha comparing multiple passages (e.g., grade 1 vs. grade 2 passages; grade 2 vs. grade 3 passages). The *t*-tests were conducted using separate groups of children reading at first-, second-, third- and fourth-grade levels. In cases in which there was not a statistically significant drop in WCPM, we modified the vocabulary and sentence structure of the passages in question in order to adjust the level of difficulty. We used lexile measures and word frequency indices as guides in making these adjustments.

During pilot testing we also conducted additional analyses of the PALS español Spelling task in order to test the accuracy of the nine-feature hierarchy and the relative difficulty of the three distinct sets of words. For these analyses, we examined data collected from 321 students in Grades 1–3 from 12 schools in 7 school districts across the U.S. All students were in dual language or transitional bilingual programs in which they were receiving literacy instruction in Spanish. Each student was administered one of two 40-word spelling inventories representing the nine spelling features identified above. We then used type-token analysis to examine the results.

Type-token analysis is a form of analysis frequently used to quantify linguistic data. By definition, types represent the number of unique occurrences of a linguistic feature, and tokens represent the total number of occurrences. In our study, tokens were calculated as the total number of times a particular spelling feature (e.g., consonant blends, diphthongs, closed syllables) appeared in the spelling inventory times the number of student attempts to spell that feature. Types were calculated as the total number of times the feature was spelled correctly, across all words and students.

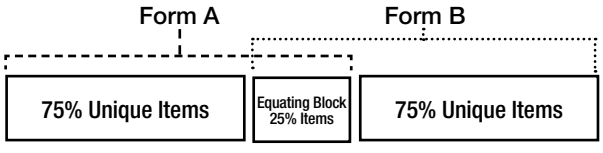
The results of this analysis revealed a clear hierarchy of orthographic features. Type-token ratios ranged from .98 for the easiest feature (open syllables) to .11 for the most difficult (affixes/roots). Our findings also supported our hypothesis that features that can be spelled relying on sound alone (Set 1) are the easiest to spell, followed by features that require an understanding of contextual constraints (Set 2) and then features that require morphological knowledge (Set 3). Mean type-token ratios were .94 for Set 1 words, .45 for Set 2, and .11 for Set 3 (see Table 3).

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 1–3. 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, the Letter Sounds task, or the Oral Reading in Context 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 on page 12.

Field Test Data Collection

Our own trained assessors administered approximately 5% of the assessments over the two years of field testing. The remaining 95% 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 1–3 was administered to 882 Spanish-speaking students in Grades 1–3 in 30 schools in Virginia, Minnesota, California, Missouri, Pennsylvania, North Carolina, Wisconsin, and Washington D.C. After revising or eliminating poorly performing items, we administered Form A again in spring 2012 to an additional 557 students in 14 schools in Virginia, Minnesota, Missouri, North Carolina, 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 1–3.

The tentative Form B of PALS español 1–3 was administered in fall 2012 to 477 Spanish-speaking students in Grades 1–3 in 7 schools in Virginia, Rhode Island, Minnesota, Wisconsin, Missouri, and Washington D.C. We then revised or eliminated poorly performing items before administering a revised version of tentative Form B to 682 additional students in spring 2013. The students in the spring 2013 field test were from 9 schools in Virginia, Illinois, Rhode Island, Georgia, 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 1–3. Table 4 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.

Table 3 Opportunities, Correct Attempts, and Type-Token Ratios (*n* = 321)

Feature	Opportunities	Correct Attempts	T-T Ratios (Feature)	T-T Ratios (Set)
Set 1 (Sound-Based)				
Open syllables	1605	1573	.98	
Closed syllables	3210	3114	.97	
Blends	1605	1547	.96	
Nasals	1605	1502	.94	
Diphthongs	1605	1362	.85	.94
Set 2 (Pattern-Based)				
Inconsistent Consonants	1605	949	.59	
Silent H	1603	617	.38	
Rule-based Consonants	1605	605	.38	.45
Set 3 (Meaning-Based)				
Affixes/Roots	1602	175	.11	.11

The analyses we used after each field testing window to identify poorly performing items were the same analyses used for this purpose during pilot testing, that is, Cronbach's alpha, D indices, and point-biserial correlations.

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.

Table 4 Field Test Sample Demographics

Demographic Category		Fall 2011 *n* = 882	Spring 2012 *n* = 557	Fall 2012 *n* = 477	Spring 2013 *n* = 682
Gender	Male	.50	.49	.49	.47
	Female	.50	.51	.51	.53
Grade	First Grade	.62	.50	.31	.35
	Second Grade	.19	.32	.31	.42
	Third Grade	.19	.18	.38	.23
FRPL	0% – 25%	.13	.19	.14	.12
	26% – 50%	.09	.02	< .01	.28
	51% – 75%	.57	.50	.83	.44
	76% – 100%	.21	.29	.03	.16
Instructional Program	ESL	.16	.13	.05	.04
	Transitional Bilingual	.10	.06	.00	.00
	Dual Language	.42	.51	.46	.63
	Spanish Immersion	.32	.30	.49	.33

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

Section III

Establishing Summed Score Criteria and Benchmarks

Benchmarks for PALS español 1–3 were developed using data gathered from 2,598 administrations of the two forms of the assessment during field testing in 2011–2012 and 2012–2013. The benchmarks for Level A (Oral Reading in Context) and for the Entry Level Word Recognition in Isolation task are theoretically driven and supported by a large body of research examining the interaction between automatic reading of words in isolation, accuracy of word reading in context, reading comprehension, and reading rate (Morris et al., 2011). The theoretical framework is based on functional reading levels. The *independent level* is the level at which students perform independently without instructional assistance and make very few errors. The *instructional level* represents a student's zone of proximal development (Vygotsky, 1978), or the level at which he/she can perform successfully and move forward with instructional scaffolding. The *frustration level* is the level at which students cannot perform successfully, even with assistance, and very little learning occurs. Benchmarks for Entry Level and Level A were set with the construct of instructional reading level in mind. For the Word Recognition in Isolation task, the instructional level is set at 75% correct word reading (Stauffer, Abrams, & Pikulski, 1978). For the Oral Reading in Context task, the instructional level corresponds to 90% to 97% accuracy in reading words in context (Afflerbach, 2007; McKenna & Stahl, 2003; Morris et al., 2011).

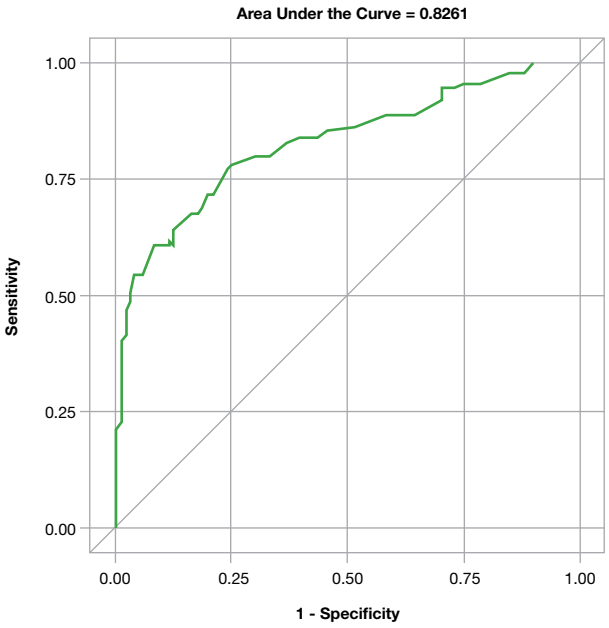
To derive benchmarks for the Level B and Level C tasks and the Entry Level Spelling task, we began by examining task means and standard deviations for scores from all students above the bottom quartile. We established initial benchmarks by subtracting one

Table 5 ROC Curve Analysis Studies

Assessment Window(s)	Grade(s)[a]	Type	*n*	AUC	Discrimination[b]
Spring 2013 (PALS esp) → Spring 2013 (L-RT)	1	Concurrent	105	.86	Excellent
Spring 2013 (PALS esp) → Spring 2013 (L-RT)	2	Concurrent	89	.93	Excellent
Spring 2013 (PALS esp) → Spring 2013 (L-RT)	3	Concurrent	81	.75	Acceptable
Spring 2012 (PALS esp) → Spring 2013 (L-RT)	1 to 2	Predictive	62	.93	Excellent
Spring 2012 (PALS esp) → Spring 2013 (L-RT)	2 to 3	Predictive	50	.81	Excellent
Fall 2011 (PALS esp) → Spring 2013 (L-RT)	1 to 2	Predictive	61	.85	Excellent
Fall 2011 (PALS esp) → Spring 2013 (L-RT)	2 to 3	Predictive	50	.75	Acceptable
Fall 2012 (PALS esp) → Fall 2012 (MAP)	1–3	Concurrent	193	.71	Acceptable
Spring 2013 (PALS esp) → Spring 2013 (MAP)	1–3	Concurrent	173	.83	Excellent
Fall 2012 (PALS esp) → Spring 2013 (MAP)	1–3	Predictive	170	.77	Acceptable

Note. L-RT = Logramos Reading Total. MAP = Measures of Academic Progress for Primary Grades. The definition of risk using the outcome variable is those students at or below the 20% national percentile rank on the MAP or Logramos.
[a]Students in Grades 1, 2, and 3 were analyzed separately for Logramos analyses but were grouped together for MAP analyses. [b]Based on Hosmer & Lemeshow (1989).

Figure 1 Area Under the Curve (AUC)



PALS español 1–3 (spring) predicting risk status using MAP (spring).

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.

Benchmarks are also provided for the PALS español 1–3 Entry Level Summed Score and Level B Summed Score. The Entry Level Summed Score benchmark is the sum of the benchmarks for the two Entry Level tasks (i.e., Word Recognition in Isolation and Spelling). The Level B Summed Score benchmark is the sum of the benchmarks for the three Level B tasks (i.e., Alphabet and Digraph Recognition, Letter Sounds, and Concept of Word).

To test the accuracy of the PALS español 1–3 Entry Level Summed Score in identifying students at risk for developing reading difficulties, we conducted 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 guidelines suggested by Hosmer and Lemeshow (1989), PALS español 1–3 has excellent discriminating capabilities. Table 5 reports AUC statistics from studies using PALS español 1–3 and two external indicators of risk: Logramos (2006), a norm-referenced reading achievement test in Spanish, and Measures of Academic Progress (MAP) for Primary Grades (2012), a norm-referenced achievement test that includes reading in English. Note that the AUC values range from .75 to .93 for the Spanish reading assessment (Logramos) and from .71 to .83 for the English (MAP) reading assessment. Figure 1 shows the AUC for spring 2013 PALS español predicting risk status using spring 2013 MAP.

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 1–3: 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 for all Entry Level (i.e., Spelling and Word Recognition in Isolation), Level B (i.e., Alphabet and Digraph Recognition, Letter Sounds, and Concept of Word in Text), and Level C (i.e., Blending and Sound-to-Letter) tasks on PALS español 1–3. All tasks were found to have good internal consistency.

Spelling. Research on how children learn to read and spell words in an alphabetic orthography has 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). Analyzing students' spelling attempts provides 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). Analyses of internal consistency based on data from field testing the final forms of this task yielded Cronbach's alphas of .94 for Form A and .93 for Form B. More information on the development of the PALS español Spelling task is provided under Construct Validity at the end of this section (Section IV).

Word Recognition in Isolation. The capacity to obtain meaning from print depends on accurate, automatic recognition of core reading vocabulary at each grade level (Snow et al., 1998). PALS español 1–3 provides five benchmark word lists to gauge students' progress: preprimer (pre-1), primer (1.1), first grade (1.2), second grade (2.2), and third grade (3.2). Analyses of internal consistency of the five benchmark word lists that make up the Word Recognition in Isolation task yielded Cronbach's alphas ranging from .91 to .95 for the five Form A word lists and .89 to .95 for the five Form B word lists.

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 Alphabet and Digraph 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 .88 for Form A and .86 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). Analyses of internal consistency yielded Cronbach's alphas of .85 for Form A and .83 for Form B of the Letter Sound Awareness task.

Table 6 Entry Level Task Reliability (Cronbach's Alpha) by Instructional Program Type

Testing Window	Grade Level	ESL	Bilingual	Dual Language	Spanish Immersion
Form A					
Fall 2011					
	1	.96 (79)	.97 (89)	.96 (148)	.96 (175)
	2	.86 (29)	N/A (0)	.92 (68)	.96 (62)
	3	.90 (22)	N/A (0)	.93 (112)	.93 (27)
Spring 2012					
	1	.97 (24)	.95 (35)	.95 (136)	.96 (85)
	2	.80 (23)	N/A (0)	.93 (94)	.95 (58)
	3	.95 (22)	N/A (0)	.94 (55)	.90 (23)
Form B					
Fall 2012					
	1	.98 (12)	N/A (0)	.94 (23)	.93 (56)
	2	N/A (0)	N/A (0)	N/A (0)	N/A (0)
	3	N/A (0)	N/A (0)	.91 (116)	.95 (61)
Spring 2013					
	1	.97 (23)	N/A (0)	.94 (125)	.96 (91)
	2	N/A (3)	N/A (0)	.92 (204)	.96 (77)
	3	N/A (0)	N/A (0)	.93 (98)	.94 (60)

Note. Subsample *n* in parentheses.

Concept of Word. Concept of word, which refers to the emergent reader's ability to match spoken words to written words as he/she recites familiar text, is an important precursor to the ability to decode text. Analyses of internal consistency of the Concept of Word task yielded Cronbach's alphas of .88 for Form A and .92 for Form B.

Blending. Phonemic awareness refers to the ability to pay attention to, identify, and manipulate phonemic segments in speech-sound units that roughly correspond to an alphabetic orthography. This awareness develops gradually over time and has a reciprocal relationship to reading (Durgunoglu & Oney, 1999; Ehri et al., 2001; Wagner, Torgesen, Laughon, Simmons, & Rashotte, 1993). Analyses of internal consistency yielded Cronbach's alphas of .76 for Form A and .81 for Form B of the Blending task.

Sound-to-Letter. The segmenting task on PALS español 1–3 assesses two kinds of knowledge necessary for learning to read: (a) speech analysis at the level of the phoneme and (b) the ability to concretize phonemic awareness and apply it to an alphabetic code. Analyses of internal consistency yielded Cronbach's alphas of .85 for Form A and .91 for Form B of the Sound-to-Letter task.

Subtask Reliability by Instructional Program

We further explored the reliability of PALS español 1–3 by using Cronbach's alpha to measure the

internal consistency of the Entry Level tasks for each grade level 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 1–3 is a reliable measure of early literacy skills for students being taught in Spanish, in English, or in both languages simultaneously (see Table 6).

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 1–3, a sub-sample of students was selected for a second administration of PALS español 1–3 after an interval of between one and two weeks. Pearson product moment correlation analysis based on the PALS español 1–3 Entry Level Summed Score yielded a correlation coefficient of .89 (n = 88).

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 1–3, 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 .83 to .98 (see Table 7).

Validity

Validity refers to how well an assessment measures what it is intended to measure. We assessed three types of validity for PALS español 1–3, 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). The content assessed in PALS español 1–3 is reading. Reading can be defined as recognition of written words that is sufficiently automatic and accurate to result in comprehension of text. Word knowledge and practice allow fluent readers to recognize words automatically and to group them into meaningful phrases. As children's reading experiences widen and their knowledge of vocabulary and word patterns expands, there is a gradual but continuous increase in word recognition and reading speed. Reading speed and fluency facilitate reading comprehension by freeing cognitive resources for interpretation (Adams, 1999; LaBerge & Samuels, 1974; Perfetti, 1985; Stanovich, 1990). Research has also established that the process of reading words, and the process of writing words,

Table 7 Inter-Rater Reliability

PALS español 1–3 Task	ICC
Word Recognition in Isolation	.87 – .96 (168)
Spelling	.98 (95)
Oral Reading in Context (accuracy)	.83 – .97 (213)
Blending	.88 – .89 (68)
Sound-to-Letter	.97 – .98 (68)

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

both draw from a common core of underlying *orthographic* or *spelling* knowledge that supports both (Perfetti, 2007).

To ensure the content validity of PALS español 1–3, we took special care to select tasks identified by research as essential to reading comprehension and to select words that are appropriate for each grade level being assessed. The Entry Level tasks of Spelling and Word Recognition in Isolation represent the fundamental orthographic knowledge necessary for fluent reading in context.

The Level A task, Oral Reading in Context, provides opportunities for teachers to assess all aspects of reading fluency (i.e., accuracy, reading rate, and prosody). Accuracy is assessed by calculating the percentage of words read accurately in each reading passage. Rate is calculated as the number of words read correctly per minute (WCPM). Teachers are also provided a three-level rubric for assessing another aspect of fluency called *prosody*, which includes such elements as expression, intonation, phrasing, and attention to punctuation. To ensure that students are not focusing solely on fluency at the expense of comprehension, questions are provided to probe their understanding.

Level B Alphabetic tasks provide a straightforward assessment of alphabet knowledge. We know that 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 Alphabet and Digraph Recognition task includes all 27 letters of the alphabet, as well as the digraphs ch and ll, which are essential to reading instruction. The Letter Sounds task includes all of the letters of the alphabet, with the exception of M, which is used as an exemplar, and H, Q, and X, which are either silent or make sounds that are ambiguous or difficult to produce in isolation. PALS español 1–3 also includes a Concept of Word task that measures how well a child can use his/her knowledge of the alphabetic code to match speech with printed text.

Level C tasks assess phonemic awareness, which is the ability to identify and manipulate sound at the phoneme level. Research has demonstrated that the ability to manipulate sounds at the phoneme level predicts later reading success in Spanish (e.g., Branum-Martin et al., 2006; Bravo et al., 2006; Carrillo, 1994). The PALS español Blending task requires that children blend individual speech sounds to form a recognizable word (e.g., /c/ + /o/ + /n/ = con). The Sound-to-Letter task asks children to isolate the individual sounds in words and match them to letters (e.g., gato begins with the /g/ sound which is represented by the letter G). In creating both these tasks, we were careful to choose words that would be in the speaking vocabulary of children in the early primary grades (Martínez Martín & García, 2004). We also gave consideration to the linguistic complexity of each sound.

Additional information on the content validity of PALS español 1–3 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 1–3.

Predictive Validity. To assess the predictive validity of PALS español 1–3 we first examined the relationship between fall PALS español 1–3 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 score,(L-WA), which measures alphabet recognition, beginning sound awareness, letter-sound relationships, and word and syllable knowledge, as well as the Logramos Reading Total score (L-RT), which is derived from the Vocabulary and Reading Comprehension subtests. Fall PALS español 1–3 Summed Scores and fall PALS español 1–3 Entry Level subtask scores were all significantly correlated ($p < .05$) with L-WA and L-RT scores obtained at the end of the year. As can be seen in Table 8, correlations between fall PALS español 1–3 scores and spring Logramos scores ranged from .56 to .67. The variable most highly correlated with Logramos was Spelling.

We also examined the relationship between PALS español 1–3 and Logramos in one- and two-year intervals. Again, we used the Logramos Word Analysis (L-WA) and Reading Total scores (L-RT) scores. As shown in Table 8, all PALS español 1–3 Entry Level subtasks were significantly correlated ($p < .05$) with the L-WA and L-RT scores at both one- and two-year across-grade intervals. Correlations ranged from .58 to .71 for the one-year interval, and from .37 to .56 for the two-year interval.

The variable most highly correlated with Logramos for the one-year predictive validity studies was the Summed Score, which includes both Spelling and Word Recognition. The variable most highly correlated with Logramos in the two-year predictive validity studies was Spelling (see Table 8).

Finally, because early literacy skills are known to transfer between languages (Durgunoglu et al., 1993), we explored the relationship between PALS español 1–3 and MAP, a literacy assessment administered in English. The fall PALS español 1–3 Summed Score was found to be significantly correlated ($p < .05$) with MAP scores administered the following spring. As can be seen in the right-hand column of Table 8, the correlation between the PALS español 1–3 Summed Score and MAP was .78.

Table 8 Predictive Validity Studies

PALS español Tasks/Scores	Logramos		MAP
	L-WA	L-RT	
Fall to Spring (Same Year)			
Summed Score	.65 (183)	.59 (264)	.78 (170)
Spelling	.67 (194)	.59 (276)	
WRI	.59 (184)	.56 (265)	
One-Year Interval			
Summed Score	.71 (63)	.62 (113)	
Spelling	.68 (63)	.58 (113)	
WRI	.66 (63)	.62 (113)	
Two-Year Interval			
Summed Score	.41 (30)	.37 (82)	
Spelling	.56 (61)	.51 (113)	
WRI	.46 (30)	.38 (82)	

Note. Summed Score = Entry Level Summed Score. WRI = Word Recognition in Isolation. L-WA = Logramos Word Analysis. L-RT = Logramos Reading Total. MAP = Measures of Academic Progress for Primary Grades. Sample n in parentheses.
All correlations are statistically significant ($p < .05$).

Concurrent Validity. To assess the concurrent validity of PALS español 1–3, we examined the relationship between the Entry Level scores on PALS español 1–3 and scores on tasks from two other assessments administered at approximately the same time: Logramos Word Analysis (L-WA), Logramos Reading Total (L-RT), and MAP. As can be seen in Table 9, the PALS español 1–3 Summed Score and both subtask scores were significantly correlated ($p < .05$) with both assessments. The highest correlations were found between the overall Summed Score and L-WA ($r = .73$) and between the overall Summed Score and MAP ($r = .71$).

We also tested the diagnostic accuracy (concurrent and predictive) of PALS español 1–3 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 1–3 data collected in the same testing window and in the previous testing window to evaluate the classification accuracy of PALS español 1–3, 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 1–3 Entry Level Summed Score. Classification accuracy for concurrent predictions ranged from 77% to 86% ($n = 375$), and classification accuracy for future predictions (i.e., from fall to spring) ranged from 75% to 80% ($n = 265$). The results from these analyses suggest that PALS español 1–3 aligns with teacher judgment of students' relative standing in reading achievement.

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 1–3 is based on a theoretical model that defines reading and writing as the application of sound, pattern, and meaning. Children must develop the ability to recognize and manipulate the sounds of the language, as well as an understanding about how orthographic patterns work to represent sound and meaning as prerequisites for reading with understanding and writing with coherence. PALS español

Table 9 Concurrent Validity Studies

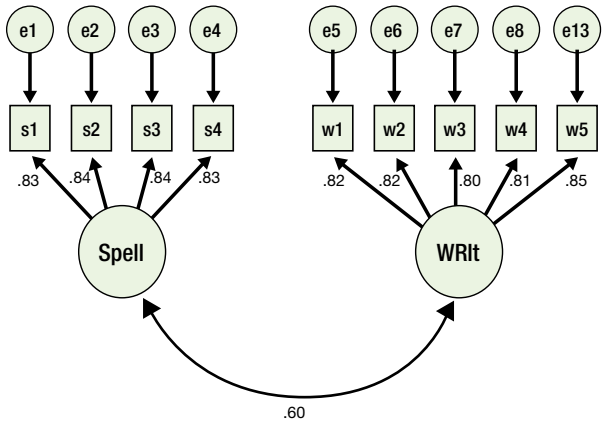
PALS español Tasks/Scores	Logramos		MAP
	L-WA	L-RT	
Fall			
Summed Score		.58 (67)	
Spelling		.51 (67)	
WRI		.59 (68)	
Spring			
Summed Score	.73 (197)	.67 (280)	.71 (173)
Spelling	.70 (197)	.64 (280)	.70 (173)
WRI	.70 (198)	.66 (281)	.66 (173)

Note. Summed Score = Entry Level Summed Score. WRI = Word Recognition in Isolation. L-WA = Logramos Word Analysis. L-RT = Logramos Reading Total. MAP = Measures of Academic Progress for Primary Grades. Sample n in parentheses.
All correlations are statistically significant ($p < .05$).

is designed to assess children's orthographic knowledge, as well as the application of that knowledge in reading and writing. Word reading and spelling are two dimensions of orthographic knowledge that undergird fluent reading, writing, and comprehension (Perfetti, 2007). 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 1–3 using data from the 2012–2013 field test (n = 220). The purpose of the study was to determine which factor structure best represented the PALS 1–3 data (i.e., a single orthographic factor model or a two-correlated orthographic factor model) and if the PALS 1–3 measured constructs would function similarly for children of different genders. We tested a one-factor model using confirmatory factor analysis (CFA) together with all the item parcels. (We converted the items from the Entry Level tasks, Word Recognition and Spelling, into parcels, or testlets, to reduce the number of parameters to be estimated and lessen the chance of Type I errors). The resulting fit indices (see Table 10) indicated that the one-factor model did not fit the data well (e.g., CFI = .75, RMSEA = .24). A succeeding two-correlated factor model was tested (see Table 10 and Figure 2), with the Spelling parcels loading on a spelling factor and the Word Recogniton in Isolation (WRI) parcels loading on a WRI factor. Model B showed a large improvement over model A and had acceptable fit indices (e.g., CFI = 1.00, RMSEA = .03, nonsignificant χ^2). This comparison allowed for an evaluation of the orthographic

Figure 2 Two-Factor Model



PALS español 1–3 two-factor model configuration and standardized factor loadings using the combined sample (n = 220). All loadings and correlations are statistically significant.

dimensionality of PALS español 1–3. The two-correlated-factor model was a significant improvement over the one-factor model and indicated a two-factor model (Word Recognition and Spelling) best represented the factor structure for PALS español 1–3 and supports the construct validity of the instrument.

Based on the finding that a two-factor model best represented the data, a multigroup confirmatory factor analysis (MGCFA) was conducted. Results from model C indicated that running the models separately for boys and girls with all loadings freely estimated had a good model fit (i.e., nonsignificant χ2, CFI = .99). The next step added equality constraints with the factor loadings of boys and girls and also constrained the correlation between the two factors to be equal. The model χ2 was still statistically significant and a χ2 difference test between model C and D showed a

Table 10 Goodness of Fit Statistics for all Models Tested

Model	Description	χ^2	df	GFI	CFI	RMSEA	AIC
A	One factor	355.38*	27	0.65	0.75	0.24	391
B	Two factor	31.05	26	0.97	1.00	0.03	69
Multigroup models for gender							
C	Freely estimated	60.25	52	0.95	0.99	0.03	136
D	Equality constrained	66.77	60	0.94	1.00	0.02	127

Table 11 PALS español 1–3 Two-Factor Item Parcel Standardized Factor Loadings

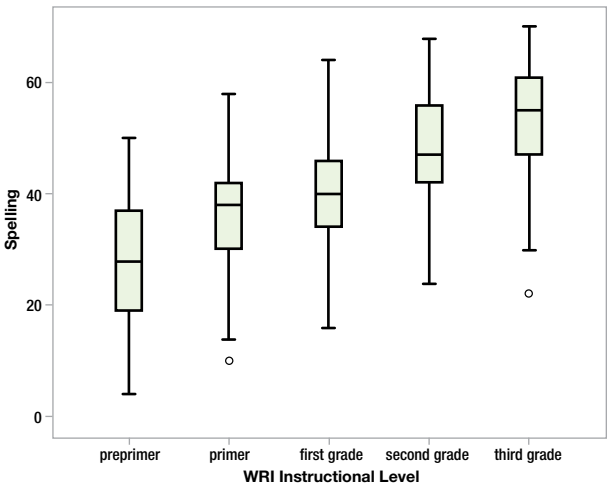
PALS Factor	Item Parcel	Combined (n = 220)	Girls (n = 103)	Boys (n = 117)
Spelling	s1	.83	.90	.80
	s2	.84	.89	.78
	s3	.84	.86	.89
	s4	.84	.94	.74
Word Recognition in Isolation	w1	.82	.89	.81
	w2	.83	.90	.75
	w3	.80	.83	.81
	w4	.81	.82	.82
	w5	.85	.88	.83
Correlation of spelling and WRI		.60	.42	.64

nonsignificant change, χ_D^2(8) = 6.52, p = .59, indicating that factor structure invariance was supported. As a result, no measurement differences were found when comparisons are made between boys and girls. All factor loadings for the combined sample, boys, and girls are shown in Table 11.

Intercorrelations. We conducted Pearson product moment correlation analysis to examine the relationships between PALS español 1–3 tasks and the overall Summed Score. The tasks that had consistently high correlations across all grade levels were Spelling and Word Recognition in Isolation (r = .66 – .79) and Spelling and Concept of Word (r = .54 – .78). Word Recognition in Isolation also had medium to high correlations with Concept of Word (r = .65 – .83), Sound-to-Letter (r = .47 – .82), and Reading Accuracy (r = .32 – .43) across all grade levels.

We used additional correlation analysis to further examine the relationship between the Word Recognition in Isolation and Spelling tasks. Research in English has demonstrated a strong relationship between word reading and spelling (Henderson, 1986). We used data from the 2012–2013 PALS español field test to explore this relationship in Spanish. We found a high correlation (r = .71, p < .001) between instructional level on the Word Recognition task (i.e., 75% accuracy) and the total score on the Spelling task (n = 550), which supports the model of developmental spelling and reading that were the basis for these tasks. Figure 3 illustrates this relationship.

Figure 3 Correlation Between Spelling and Word Recognition in Isolation Instructional Level



Section V

Summary

The technical adequacy of PALS español 1–3 has been established through 2,598 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 .76 to .95, demonstrating good internal consistency. Inter-rater reliabilities expressed as intraclass correlation coefficients ranged from .83 to .98, indicating that PALS español 1–3 can be scored consistently across individuals.

Analyses of field test data also support the content, construct, and criterion-related validity of PALS español 1–3. 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 1–3 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 1–3 is an assessment tool with good evidence of reliability and validity that can be used effectively to screen Spanish-speaking students in Grades 1–3 for possible reading difficulties. PALS español 1–3 shows evidence of internal consistency, demonstrating that the items within each task produce similar results in measuring the same construct. PALS español 1–3 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 1–3 also demonstrates good evidence of content, construct, and criterion-related validity, suggesting that PALS español 1–3 truly measures the underlying constructs associated with literacy development in Spanish.

Section VI
References

Adams, M. (1990). *Beginning to Read: Thinking and Learning about Print.* Cambridge, MA: MIT Press.

Afflerbach, P. (2007). *Understanding and using reading assessment (K–12).* Newark, DE: International Reading Association.

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.

Barr, R., Blachoqica, C., & Wogman-Sadow, M. (1995). *Reading diagnosis for teachers: An instructional approach* (3rd ed.). White Plains, NY: Longman.

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., & Francis, D. J. (2006). Bilingual phonological awareness: Multilevel construct validation among Spanish-speaking kindergarteners in transitional bilingual education classrooms. *Journal of Educational Psychology, 98*(1), 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.

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

Davies, M. (2002). *Corpus del español: 100 million words, 1200s–1900s.* Available online at http://www.corpusdelespanol.org.

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.

Defior, S., Jiménez- Fernández, G., & Serrano, F. (2009). Complexity and lexicality effects on the acquisition of Spanish spelling. *Learning and Instruction, 19*, 55–65.

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.

Durgunoglu, A. Y., & Oney, B. (1999). A cross-linguistic comparison of phonological awareness and word recognition. *Reading and Writing: An Interdisciplinary Journal, 11*, 281–299.

Ehri, L. C., Nunes, S. R., Willows, D. M., Schuster, B., Yaghoub-Zadeh, Z., & Shanahan, T. (2001). Phonemic awareness instruction helps children learn to read: Evidence from the National Reading Panel's meta-analysis. *Reading Research Quarterly, 36*, 250–287.

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.

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. (1986). A developmental perspective of formal spelling instruction through alphabet, pattern, and meaning. *The Elementary School Journal, 86*, 304–316.

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.

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.

Johnson, M. S., Kress, R. A., & Pikulski, J. J. (1987). *Informal reading inventories* (2nd ed.). Newark, DE: International Reading Association.

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

LaBerge, D., & Samuels, S.J. (1974). Toward a theory of automatic information processing in reading. *Cognitive Psychology, 6*, 293–323.

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.

Martínez Martín, J. A., & García, E. (2004). *Diccionario: Frecuencias del castellano escrito en niños de 6 a 12 años.* Salamanca: Publicaciones Universidad Pontificia.

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.

McKenna, M., & Stahl, S. (2003). *Assessment for Reading Instruction.* New York: Guilford Press.

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.

Morris, D., Bloodgood, J.W., Perney, J., Frye, E.M., Kucan, L., Trathen, W., Ward, D., & Schlagal, R. (2011). Validating craft knowledge: An empirical examination of elementary-grade students' performance on an informal reading assessment. *The Elementary School Journal, 112*, 2, 205–233.

Perfetti, C. (1985). *Reading Ability.* New York: Oxford University Press.

Perfetti, C. (2007). Reading ability: Lexical quality to comprehension. Scientific Studies of Reading, 11(4), 357–383.

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.

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

Stanovich, K. (1990). Concepts in developmental theories of reading skills: Cognitive resources, automaticity, and modularity. *Developmental Review, 10*, 72–100.

, R., Abrams, J., & Pikulski, J. (1978). *Diagnosis, Correction, and Prevention of Reading Disabilities.* New York: Harper & Row.

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes.* Cambridge, MA: Harvard University Press.

Wagner, R. K., Torgesen, J. K., Laughon, P., Simmons, K., & Roshotte, C. A. (1993). Development of young readers' phonological processing abilities. *Journal of Educational Psychology, 85*, 83–103.

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