



Phonological Awareness Literacy Screening

Marcia Invernizzi • Connie Juel • Linda Swank • Joanne Meier
University of Virginia • Curry School of Education

K Technical Reference



For questions about PALS-K, please contact:
Phonological Awareness Literacy Screening (PALS)
1-888-UVA-PALS (1-888-882-7257) or (434) 982-2780
Fax: (434) 982-2793
e-mail address: pals@virginia.edu • Web site: <http://pals.virginia.edu>

©2004–2015 by The Rector and The Board of Visitors of the University of Virginia. All Rights Reserved.

Graphic Design: Branner Graphic Design

Printed in the United States of America



K Technical Reference

This document is a supplement to the *PALS-K Administration & Scoring Guide*.

Marcia Invernizzi • Connie Juel • Linda Swank • Joanne Meier
Virginia State Department of Education
University of Virginia • Curry School of Education

Acknowledgments

Development of the Phonological Awareness Literacy Screening (PALS-K) has been supported by the Virginia Department of Education through Virginia's Early Intervention Reading Initiative. Without the support provided by the Department, the test development activity required for this assessment would not be possible.

The PALS Office would like to thank Dr. Randall Robey, Dr. Timothy Landrum, and Dr. Tonya Moon at the University of Virginia for their contributions to the technical adequacy of PALS-K. Thanks go also to division representatives, principals, and teachers throughout Virginia who have participated in pilots. Thanks to their participation, the PALS office is able to ensure that classroom teachers have a literacy screening tool with good evidence of reliability and validity. We also wish to thank several generations of graduate students from the McGuffey Reading Center who have contributed to the development of PALS-K tasks.

Section I

5 Phonological Awareness Literacy Screening for Kindergarten (PALS-K)

- 5 Purposes, Uses, and Limitations
- 5 Overview
- 6 Background
- 6 Virginia's Early Intervention Reading Initiative (EIRI)
- 7 Funding for Virginia's Early Intervention Reading Initiative
- 7 Virginia's Standards of Learning (SOL) and PALS

Section II

9 Description of PALS-K

- 7 Domains
- 7 Scoring
- 8 Forms

Section III

11 Item Development and Field-Testing

- 11 Phonological Awareness Tasks
- 11 Rhyme and Beginning Sound Awareness
- 13 Literacy Tasks
- 13 Alphabet Knowledge
- 14 Letter-Sound Knowledge
- 15 Concept of Word
- 17 Word Recognition in Isolation
- 18 Feedback from the Field
- 18 Outside Review
- 19 Advisory Review Panel
- 19 External Review

Section IV

20 Establishing Summed Score Criteria and Benchmarks

- 21 Benchmarks and Discriminant Analysis (DA)

Section V

22 Technical Adequacy

- 23 Broad Representation of Students
- 23 Pilots
- 23 Summary Statistics
- 24 Reliability
- 26 Test-retest Reliability

26	Subtask Reliability
26	Inter-rater Reliability
29	Internet Data Entry Reliability
29	Validity
29	Content Validity
30	Criterion-related Validity
31	Construct Validity
35	Differential Item Functioning

Section VI

36 Summary

Section VII

37 References

Section VIII

39 Endnotes

Section I

Phonological Awareness Literacy Screening for Kindergarten (PALS-K)

In this section we

- provide an overview of the purpose and use of PALS;
- provide an overview of Virginia's Early Intervention Reading Initiative (EIRI);
- describe the way that funding is allocated for divisions that participate in the EIRI;
- show how PALS supports Virginia's Standards of Learning (SOL);
- describe briefly the PALS-K instrument.

More detailed information about the instrument is available from our website (pals.virginia.edu).

Purposes, Uses, and Limitations

The major purpose of the Phonological Awareness Literacy Screening for Kindergarten (PALS-K) is to identify students who perform below grade-level expectations in several important literacy fundamentals, and thus are at risk of reading difficulties and delays. As a diagnostic tool, PALS-K can be used to assess what students already know about the English writing system and what they need to learn to become readers. PALS-K has demonstrated good evidence of reliability and construct, concurrent, and predictive validity.

However, like any other assessment tool, PALS-K is just one means to assess a student's overall literacy competence. Other important information includes additional early literacy assessment data, parent information, the child's interest in books, and teacher judgment. While PALS-K provides reliable screening for developmental milestones in literacy acquisition, one measure of an emergent reader's performance is

never sufficient when making high-stakes decisions such as summer school placement or retention.

Overview

Consisting of three screening instruments, the Phonological Awareness Literacy Screening (PALSPreK, PALS-K, and PALS Plus for grades 1–8), measures young children's knowledge of important literacy fundamentals, including:

- phonological awareness;
- alphabet knowledge;
- knowledge of letter sounds;
- spelling;
- concept of word;
- word recognition in isolation.

The major purpose of PALS is to identify students who are performing below grade-level expectations in these areas and may be in need of additional reading instruction beyond what is typically provided to developing readers. Note that meeting the Summed Score benchmark does not imply that the student is on grade level, but only that the student met the level of minimal competency necessary to benefit from typical classroom literacy instruction. A secondary and logical extension of this goal is to provide teachers with explicit information about what their students know of these literacy fundamentals so that they can more effectively tailor their teaching to their students' needs.

This Technical Reference includes a description of the background and rationale underlying PALS-K, the process of task and item development and field-testing, and the technical adequacy of the

instrument (validity and reliability). In preparing this Technical Reference, we followed the Standards for Educational and Psychological Testing (1999), prepared jointly by the American Educational Research Association (AERA), the American Psychological Association (APA), and the National Council on Measurement in Education (NCME). Explicit instructions for the administration and scoring of PALS-K are included in a separate PALS-K Administration and Scoring Guide. The results for the statewide screening for each cohort are available in separate annual reports.

Background

The Phonological Awareness Literacy Screening for Kindergarten (PALS-K) is the state-provided screening tool for the Virginia Early Intervention Reading Initiative (EIRI), and is designed for use in kindergarten. The purpose of the EIRI is to reduce the number of children with reading problems through early detection and to accelerate their learning of research-identified emergent and early literacy skills.

Virginia's Early Intervention Reading Initiative (EIRI)

The 1997 Virginia Acts of Assembly, Chapter 924, Item 140, initially established the Early Intervention Reading Initiative. The state initiative allocated funds to help participating school divisions identify children in need of additional instruction and to provide early intervention services to students with diagnosed needs. Participating school divisions were allowed to implement the initiative in either kindergarten or first grade.

In the 2000-01 legislative session, the Governor and the General Assembly provided funding to expand the EIRI to third grade. Participating school divisions are now required to screen students in kindergarten through third grade either with a diagnostic assessment approved by the Virginia Department of Education or with PALS, the state-provided instru-

ment. Many of the same conditions from the earlier initiative apply:

- All students in kindergarten through second grade must be screened annually;
- All students not meeting the benchmark for their grade level must receive, in addition to regular classroom instruction, intervention services;
- All students in kindergarten through second grade who receive intervention services must be assessed again during the first screening period following the intervention. (Note that third-grade students are only screened in the fall if they are new to Virginia schools, or if they received intervention services over the summer; spring screening for third-graders is optional);
- All screening results must be reported to the PALS Office at the University of Virginia via the PALS website (pals.virginia.edu).

In 2002, the Virginia Department of Education changed the screening period for the EIRI from fall to spring. Also, a high benchmark was added for first- and second-grade students clearly performing above grade-level expectations. Students attaining this high benchmark would no longer need to be screened for the EIRI. These changes enhance the EIRI by:

- allowing intervention services for all students in first, second, and third grades to start at the beginning of the school year or during the summer;
- eliminating the problem created by fall screening for year-round schools and schools that start before Labor Day;
- allowing Title I to use PALS as their screening instrument for reading services, thereby eliminating the use of a second screening;
- reducing the amount of time required for screening.

An EIRI timeline for PALS screening is shown in Table 1.

Section II

Description of PALS-K

In this section we briefly describe the parts of PALS-K. Table 3 outlines the conceptual framework for the instrument.

Among the most effective strategies for preventing reading problems is first to identify early and accurately children who are experiencing difficulties in acquiring fundamental skills, and second, to ensure that these children attain critical beginning literacy skills through additional instruction. This approach can be viewed as simultaneously proactive and preventative.

A substantial research base suggests key variables that help identify children most likely to experience subsequent difficulties with reading achievement.³ This research indicates that measures of phonological awareness, alphabet knowledge, letter-sound knowledge, and other elements of early literacy (e.g., phonetic spelling, word recognition) serve as robust predictors of children's later literacy achievement.

Domains

PALS-K measures kindergarten students' development in each of these dimensions related to early literacy preparation, most notably in awareness of speech sounds and knowledge of print. The phonological awareness component of the PALS-K instrument is a measure of young children's ability to identify rhyme units and isolate beginning sounds. The literacy component of the PALS-K instrument is a measure of young children's knowledge of important literacy fundamentals:

- alphabet knowledge;
- knowledge of letter sounds;
- phoneme-grapheme correspondences;
- concept of word;
- word recognition.

Table 3 highlights the conceptual framework and lists the subtasks that make up the PALS-K assessment tool.

Scoring

Students' scores on selected PALS-K subtasks are added together to create a Summed Score. This score is compared against developmental expectations for fall and for spring. Students with Summed Scores below expectations are provided additional instruction funded by Virginia's Early Intervention Reading Initiative.

Students demonstrate their skill in each domain to their classroom teacher, who administers PALS-K in the classroom (after reading the PALS-K Administration and Scoring Guide). The tasks do not have a time limit and are tested one-on-one, except for the Group Rhyme Awareness, Group Beginning

Table 3 Conceptual Framework for PALS-K

Domain	Task
Phonological Awareness	Rhyme Awareness
	Beginning Sound Awareness
Literacy Skills	Alphabet Knowledge
	Letter Sounds
	Spelling
	Concept of Word
	Word Recognition in Isolation

Sound Awareness, and Spelling tasks, which can be administered in small groups. A criterion score or benchmark is provided for each task to help measure a minimal level of competency in that domain.

Procedures for PALS-K administration and specifications for scoring may be found in the PALS-K Administration and Scoring Guide. A description of how the criterion scores or benchmarks were established may be found later in this Technical Reference. The following section contains a detailed description of how PALS items and tasks were developed and field-tested.

Forms

Three forms of PALS-K are now in use. Forms A and B are used in alternate years: Form A was used in 2013–14, Form B was used in 2014–15. Form C is the optional mid-year form.

Section III

Item Development and Field-Testing

In this section we describe the various tasks of PALS-K:

- Rhyme and beginning sound awareness;
- Alphabet knowledge;
- Letter-sound awareness;
- Letter sounds;
- Spelling;
- Concept of word;
- Word recognition in isolation.

We also describe feedback we receive from experts in the field.

PALS-K evolved from the McGuffey Reading Center's Test of Early Word Knowledge (EWK), which later became McGuffey's Assessment of Literacy Acquisition (ALA). Both of these early literacy assessment procedures have been adapted, expanded, and applied in early intervention settings across the country, most notably by Darrell Morris. Morris' Early Reading Screening Inventory (ERSI) (see Perney, Morris, & Carter, 1997) has been used extensively across the country and includes many of the same tasks contained in PALS-K.

The tasks presented in PALS-K are a representative sample of tasks found in other measures of early literacy. Items were selected because of their previous history in phonological awareness and early literacy research, and because of their correlation with Virginia's Standards of Learning. Item selection and field-testing procedures for the original and revised versions of PALS-K are described below.

Phonological Awareness Tasks

Phonological awareness refers to the ability to pay attention to, identify, and manipulate sound units

within spoken words. The research literature on phonological awareness identifies two skills significantly related to reading outcomes: (a) rhyme awareness, and (b) individual phoneme awareness.⁴ Items in PALS-K were selected to represent these two categories of sound awareness and to meet three attributes of measurement. First, the items selected needed to be of moderate difficulty for kindergarten children. Second, the items selected needed to have strong predictive relationships to reading outcomes. Measures of rhyme awareness and phonemic awareness are well documented as predictive of reading outcome.⁵ Third, the selected items needed to be adaptable to group assessment procedures. Because the format for both tasks subsumed under Phonological Awareness (Rhyme Awareness and Beginning Sound Awareness) is similar, the following section describes the development of these tasks concurrently.

Rhyme and Beginning Sound Awareness

New format. Traditional measures of phonological awareness typically assess students in an individual format, using oral assessment procedures. In this way, obtaining phonological awareness data on an entire class can become a lengthy and time-consuming process. The items on the PALS-K Group Rhyme Awareness and Group Beginning Sound Awareness tasks allow teachers to assess students in small groups of five or fewer. Only those students who exhibit difficulty in the group screening require individual follow-up to gather more detailed information about which sound units present difficulty for a given student.

Picture prompts. We selected developmentally appropriate pictures with a prior history in phonological awareness research. Items selected met two criteria: (a) stimuli had been used previously with preschool and primary-age children to assess pho-

nological awareness, thus establishing predictive outcomes; and (b) pictures were easily recognizable and represented age-appropriate vocabulary. The first criterion was met by selecting stimulus words from past prediction studies.⁶ We met the second criterion by selecting picture templates previously used successfully with preschool and primary-age children, and by having an artist draw similar renderings of pictures. The pictures represent one-syllable, high-frequency words appropriate for kindergarten children.⁷ We included only single-syllable words with concrete meanings that could be represented pictorially.

Field review. The PALS-K pictures and teacher administration instructions were reviewed by a panel of primary classroom teachers, elementary administrators, university researchers, and Virginia Department of Education personnel in assessment, early childhood, and elementary instruction. Following approval by the 15-member panel, the phonological awareness measures were then piloted with 50 kindergarten and first-grade children in two school divisions in different parts of the state, while classroom teachers and administrative personnel observed. Following the first administration, classroom teachers and administrative personnel were trained to re-administer the phonological awareness tasks. Within a three-week period, they retested the same students for preliminary test-retest reliability data. Following the re-administration, teachers and administrators provided oral and written feedback on the instructions and on students' performance. They also provided their own reactions to the procedure and suggested changes. Their suggested changes were submitted to the 15-member panel for final approval and incorporation into PALS-K. This set of procedures resulted in the current PALS-K phonological awareness tasks, Rhyme Awareness and Beginning Sound Awareness.

Field testing. The phonological awareness items were administered to 53,425 kindergartners and first-graders in the fall of 1997 and to 65,619 kindergartners and first-graders in the fall of 1998. Four

types of picture revisions resulted from an analysis of the 1997 and 1998 samples. First, controversial pictures were changed to reflect more appropriate items. For example, the picture of the pipe in the Group Beginning Sound Awareness task was eliminated and replaced with a picture of a bus. Second, ambiguous pictures were redrawn to provide greater clarity. For example, the picture of the rock was redrawn to look more like a rock. Third, unfamiliar pictures were replaced with more common items. For example, the picture of the fountain pen was replaced with a picture of the more common ballpoint pen. Fourth, random sound relations among pictures in the same row were eliminated, so that no sound within the name of the target picture occurred in any position in any other picture within the row. For example, the picture of the tie was changed to a picture of a bell so as not to prompt attention inadvertently to the /t/ sound at the end of the target picture heart. The order of the pictures was also changed in some cases to ensure that correct responses were distributed randomly across items; thus, scores would not be biased if, for example, a child simply chose the first picture in each row.

Additional testing. Further pilot data on individual items were collected in Fall 2001 with 1,855 kindergarten children for Group Rhyme Awareness and 1,862 kindergarten children for Group Beginning Sound Awareness. In Spring 2004 data on individual items were collected from 1,417 kindergarten children for Group Rhyme Awareness and 1,227 kindergarten children for Group Beginning Sound Awareness. These phonological awareness tasks and items within these tasks were examined with regard to (a) item-to-total correlations, (b) Cronbach's alpha (an index of internal consistency based on the average correlation of items within a task),⁸ and (c) item means (level of difficulty). Items were considered for removal if they had low item-to-total correlations, were too easy or difficult (i.e., nearly all students responded correctly or nearly all students missed the item), or if scales yielded alpha coefficients less than .80. In these pilot samples, item-to-total correlations for each item were moderate

to high, ranging from .37 to .70. Alpha coefficients for Group Rhyme Awareness and Group Beginning Sound Awareness were high, ranging from .83 to .87 across samples. Means for each item indicated that all items were of acceptable difficulty. Based on these results, no items in the phonological awareness section were replaced.

An additional pilot test was conducted in January 2005 with 193 kindergarten students. This pilot established the internal consistency of the Form C Group and Individual Beginning Sound tasks (alpha coefficients of .87 and .94, respectively), and the Group and Individual Rhyme tasks (alpha coefficients of .85 and .88, respectively).

Individual testing. Students who do not meet the benchmark on the group phonological awareness tasks (Group Rhyme Awareness, Group Beginning Sound Awareness) are administered the individual versions of these tasks (Individual Rhyme Awareness, Individual Beginning Sound Awareness). The individual scores are included in the student's Summed Score.

Analyses of PALS data show that most students (70% for beginning sound, 88% for rhyme) perform better on the individually administered tasks. The students who do not perform better under the individual condition generally are those who scored below benchmarks on other tasks as well. In fact, further analysis in Fall 2002 showed that these lower individual scores affected the identification status for only 0.2% (in the case of Individual Rhyme Awareness) or 0.3% (in the case of Individual Beginning Sound Awareness) of kindergarten students screened. All students in these analyses scored below benchmark on both group and individual tasks, so the effect of this phenomenon on the appropriate identification of students needing additional reading instruction was negligible.

Literacy Tasks

The items for the literacy screening component of PALS-K are similar, if not identical, to many of the items of the ERSI⁹ and the Book Buddies Early Literacy Screening (BBELS).¹⁰ Some items within the Alphabet Knowledge, Letter-Sound Knowledge, Word Recognition in Isolation, and Concept of Word sections of PALS-K are common to all three instruments. These tasks have been used for a number of years with thousands of kindergarten and first-grade children in central Virginia; with thousands of first graders in North Carolina, Illinois, Montana, and Tennessee; and in at least 25 sites elsewhere across the country. Previous research on the ERSI and the BBELS provides support for the tasks on the literacy component of PALS-K.¹¹ Analyses of validity and reliability over eight cohorts of Virginia's EIRI, Internet surveys, and teacher feedback all contributed to the item development of the PALS-K literacy tasks.

Alphabet Knowledge

The single best predictor—on its own—of early reading achievement is accurate, rapid naming of the letters of the alphabet.¹² Children from the first PALS-K cohort were initially asked to name all of the letters of the alphabet in both upper and lower case.¹³ At that time, 52,660 kindergarten and first-grade children were administered the upper- and lower-case Alphabet Recognition tasks. Children were asked to name a series of 26 randomly presented letters, first in upper case, then again in lower case.

Item analyses from the 1997 statewide sample demonstrated ceiling effects for upper-case recognition among first graders. Since upper-case recognition and lower-case recognition were significantly and highly correlated ($r = .94$ for the kindergarten sample and .83 for first grade), and no ceiling effects occurred for lower-case letters, PALS 1998-99 was revised to include Alphabet Recognition for lower-case letters only. Teacher feedback from the 1998 administration also prompted a change in the order of letter presentation. Previously, the first alphabet

item encountered was a lower-case b, a letter frequently confused with lower-case d. On the current PALS-K, the first item encountered is an m. Inter-rater reliabilities for the Lower-Case Alphabet Recognition task have been consistently high ($r = .99, p < .01$).

Letter-Sound Knowledge

In addition to naming the letters of the alphabet, emergent 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 is difficult for young children, and requires explicit awareness of individual phonemes. PALS-K assesses both children's knowledge of letter sounds and their application of that knowledge in two tasks: Letter Sounds and Spelling.

Letter Sounds. In the Letter Sounds task, children are asked to touch each letter and say the sound it represents. Only the lax (or short) vowel sound for each vowel is scored as correct, and only the hard sound for C and G is scored as correct. Children are prompted for “the other sound” a letter makes in cases where they provide a long vowel sound or the soft sounds for C or G. Inter-rater reliabilities for the Letter Sounds task have been consistently high: $r = .98$ to $.99$ ($p < .01$). Because research has shown that kindergartners recognize more upper-case than lower-case letters, knowledge of grapheme-phoneme correspondence is assessed using upper-case letters in PALS-K. In the first cohort of the EIRI, all of the upper-case letters were used, with the exception of X and Q, since neither of these letters can be pronounced in isolation. Qu was substituted for Q and Sh took the place of X. Negative feedback from the first PALS-K administration regarding Qu prompted the elimination of this item in the 1998 edition. Ch, a more frequently occurring digraph, replaced Qu, and Th replaced M, which became the letter used as an example in the directions.

Spelling. Application of letter-sound knowledge in invented spelling tasks is an excellent predictor of word recognition in young children¹⁴ and among the

best predictors of word analysis and word synthesis.¹⁵ In the first cohort of Virginia's EIRI, 35,518 kindergarten and 16,136 first-grade students attempted to spell five consonant-vowel-consonant (CVC) words in the fall of the academic year. In the second year, 50,949 kindergartners and 14,670 first graders attempted to spell the same five high-frequency words. In both samples, children's spellings were scored for the number of phonemes represented. The Spelling task has consistently been a reliable discriminator of children in need of additional instruction in phonological awareness and early literacy skills in both kindergarten and first grade. Inter-rater reliabilities have remained high for all statewide samples: $r = .99$ ($p < .01$).

In Spring 2001, two sets of five new spelling words were piloted among 847 kindergartners in 22 different school divisions across the Commonwealth of Virginia. Then, in Fall 2001, two additional sets of five spelling words were piloted among 1,980 kindergartners in 52 different school divisions across the Commonwealth of Virginia. The piloted items were all high-frequency CVC words.

Words for the piloted spelling inventories were selected from a pool of words used in previous research in the Virginia Spelling Studies.¹⁶ Specific words were selected by frequency of occurrence and by each word's linguistic attributes. That is, words were selected to elicit responses to particular speech sounds and high-frequency CVC phonograms typically encountered in print early on. Five words were selected for each form. All pilots assessed student performance on the representation of beginning, middle, and ending speech sounds and the total number of words spelled correctly. In scoring each word, students received a point for the phonetic representation of the beginning, middle, and ending sound. Another point was awarded if the entire word was spelled correctly. In this way, students were credited for phonetic representation of individual phonemes regardless of whole-word spellings.

Individual words from all pilot lists were analyzed using the following criteria:

- teacher feedback;
- item means (level of difficulty);
- item-to-total correlations;
- Cronbach's alpha.

Words were considered for removal if they received negative feedback from more than two teachers in the pilot sample, if they were too easy or difficult, if they had low item-to-total correlations, or if a given spelling list had an alpha less than .80. None of the piloted spelling words from the Spring 2001 pilot warranted replacement based on empirical grounds. Spelling lists had alpha coefficients greater than .90; all item-to-total correlations were in the range of .49 to .72; and all piloted words were of acceptable difficulty. However, 32% of the teachers participating in the pilot study voiced concerns over the word *jog* because of perceived unfairness regarding *j* and *g* in the same word, so this word was removed. The piloted spelling lists and the original spelling lists were significantly correlated ($r = .70, p < .001$).

Spelling lists from the Fall 2001 pilot also had alpha coefficients greater than .90; all item-to-total correlations were in the range of .49 to .80; and all piloted words showed evidence of acceptable difficulty. Although both piloted lists were acceptable on these criteria, one word list was consistently superior on all criteria (e.g., higher alpha), and it was selected for use in the PALS-K materials. Teacher feedback indicated that kindergarten students were confused when sentences were provided with the spelling words; therefore, we did not include spelling sentences with the PALS-K materials.

Following selection of the spelling words from the Spring 2001 pilot, we examined students' spelling samples ($n = 354$) to determine the most common phonetic substitutions made by kindergarten students. Where students consistently represented a particular letter, we compared these phonetic representations with the developmental spelling literature in order to verify the accuracy of the spelling scoring grid.

In Spring 2002, we also piloted an expanded 12-word spelling list that was not only longer than the current PALS-K spelling list but also included a more complete analysis of spelling features. Our aim in this pilot was to determine whether we could better approximate the PALS 1–3 spelling task that first graders would encounter, and more importantly to assess whether we could further strengthen the relationship between PALS-K scores from spring of kindergarten, and PALS 1–3 scores for first graders in the fall. Our analyses of these pilot data suggested that the enhanced spelling task added little to the prediction of first grade scores based on PALS-K scores; that is, it offered little or no statistical benefit over the present Spelling task. A change in the spelling task for this purpose was not warranted based on this pilot.

A final additional list of five spelling words was piloted in Spring 2003 with 1,565 kindergarten students. Again this spelling list and individual words were examined using the same criteria as for earlier lists: item difficulty, item-to-total correlations, and Cronbach's alpha. This piloted list met all statistical criteria, correlated well with the Spring 2003 PALS-K spelling list ($r = .88, p < .001$) and received no negative teacher feedback; thus no changes were necessary.

Concept of Word

Concept of word refers to the emergent reader's ability to match spoken words to written words as s/he reads.¹⁷ Research has shown that a stable concept of word in text can facilitate a child's awareness of the individual sounds within words. Until children can point to individual words accurately within a line of text, they will be unable to learn new words while reading or to attend effectively to letter-sound cues at the beginning of words in running text. The ability to fully segment all the phonemes within words appears to follow concept of word attainment.¹⁸ Children with a solid concept of word will recognize words they didn't know prior to reading a memorized or familiar text, even when these words are presented out of context.

Development of the Concept of Word task. In 1997, 34,848 kindergarten and 3,586 first grade students were administered the Concept of Word finger-point reading task. Qualitative feedback from the field indicated that some children were unfamiliar with the content of the text used that year, which featured a farmer and a goat. Although familiarity with the story content would not have affected the outcome of the measure, the content was changed in the 1998 version of PALS to a standard nursery rhyme. Further feedback from the PALS Internet survey indicated a teacher preference for the book format of 1997 and for a more familiar nursery rhyme. As a result, PALS-K uses simple nursery rhymes presented in a book format, one line to a page. The administration and scoring of the PALS Concept of Word task remained unchanged for the first four cohorts of Virginia's EIRI.

Field testing. Teachers providing feedback from the 2000-01 school year requested another nursery rhyme to alternate from fall to spring. In response to this need, multiple nursery rhymes were field-tested with 1,405 end-of-year kindergartners and first-graders in Spring and Fall 2001. A new procedure for administration of the Concept of Word task was also piloted at the same times. The new procedure involved pre-testing words from the rhyme prior to the finger-point reading exercise. The same words were post-tested after the finger-pointing exercise, to see if any words were "picked up" in the process. Words identified in the post-test that were not known in the pre-test are words learned by virtue of participating in the Concept of Word task itself. Among pilot sample scores from Spring 2001 ($n = 276$), the post-test subscore at the end of kindergarten was found to be significantly correlated with the preprimer Word Recognition in Isolation scores that children earned the following fall (i.e., at the beginning of first grade) ($r = .79$).

From these pilots, nursery rhymes were selected for use if they received positive feedback from the pilot teachers and yielded reliability coefficients of .80 or higher. Reliability was assessed for pre- and post-test

word lists using Cronbach's alpha. Pre-test word list alphas ranged from .76 ($n = 162$) to .90 ($n = 402$) and post-test word list alphas ranged from .81 ($n = 161$) to .93 ($n = 421$). Therefore, no words needed to be replaced in the pre- and post-test word lists. Care was taken to mix the order of words in the word lists so that these lists did not reflect the order of word presentation in the rhyme itself. Selection of target words in the Word Identification portion was based on both the position of words in the sentence as well as word difficulty. For instance, in each poem some words from the beginning, middle, and end of the lines were assessed. Additional modifications were made to the test layout and to some illustrations accompanying the rhymes, based on teacher feedback.

Changes. In Spring 2003, minor changes were made to the Concept of Word task to enhance the relationship between PALS-K and PALS 1–3, particularly across the spring kindergarten to fall first grade screenings. First, the Pre-test Word List task was eliminated. This was based on both teacher feedback and statistical analyses suggesting that the pretest word list added little to the predictive validity of Concept of Word in relation to later PALS scores. Second, the Concept of Word Post-test Word List score was included in the PALS-K Summed Score beginning in Spring 2003. This decision was based on statistical analyses suggesting that the Post-test Word List significantly enhances the predictive validity of PALS-K in relation to PALS 1–3 scores. The Pointing and Word Identification subtasks in Concept of Word remained the same.

Concept of Word poems. In Fall 2003 three Concept of Word poems were piloted (Baa Baa Black Sheep, Little Teapot, and Little Turtle). In Spring 2004 two additional Concept of Word poems were piloted (Little Bo Peep and Little Boy Blue). In each case, we again examined (a) teacher feedback, (b) the internal consistency of items within subtasks, and (c) the relationship (correlation) between the piloted tasks and the regular PALS-K Concept of Word tasks administered at the same time. From the Fall pilot,

Little Turtle was chosen based on the combination of very positive teacher feedback, strong and significant correlation between the pilot word list and the PALS-K Concept of Word list ($r = .84$, $p < .01$, $n = 1,776$), and acceptable internal consistency (Cronbach's $\alpha = .84$).

From the Spring 2004 pilot, Little Bo Peep was chosen, again based on positive teacher feedback, significant correlation between the piloted COW word list and the PALS-K COW word list ($r = .76$, $p < .01$, $n = 1,280$), and acceptable internal consistency (Cronbach's $\alpha = .88$).

Word Recognition in Isolation

Word lists. Since the capacity to obtain meaning from print depends so strongly on accurate, automatic recognition of words,¹⁹ PALS-K provides three optional word lists to gauge advancing kindergartners' progress throughout the year: preprimer (pre-1), primer (1.1), and first grade (1.2). Each list represents a random sample from a database of words created from a variety of sources.

Word lists for PALS-K were developed in conjunction with the preprimer through third-grade word lists for PALS 1–3. Originally, word lists were generated from a database of words created from three of the most frequently used basal readers in Virginia. These included the Harcourt Brace Signature series and the Scott Foresman series from 1997 and 1999. Then, words from the first-, second-, and third-grade lists from the *EDL Core Vocabularies in Reading, Mathematics, Science, and Social Studies* (1997) were added to the database. The EDL Core Vocabularies provides a core reading vocabulary by grade, comprised of words derived from a survey of nine basal reading series. Words from the *100 Most Frequent Words in Books for Beginning Readers*²⁰ were added to the preprimer, primer, and first-grade word pools.

Data base expanded. After the first year, the database was expanded to include words from grade-level lists in spelling and vocabulary books. These included words from *Teaching Spelling*,²¹ *A Reason*

for Spelling,²² *A Combined Word List*,²³ and *A Basic Vocabulary of Elementary School Children*.²⁴ The database now includes all of these words, plus the words from graded word lists located in informal reading inventories and other well-known published assessments with grade-level lists. Words were added to the database from the *Qualitative Reading Inventory II* (QRI-II),²⁵ *The Stieglitz Informal Reading Inventory* (1997), the *Bader Reading and Language Inventory* (1998), the *Decoding Skills Test*,²⁶ the *Ekwall/Shanker Reading Inventory*,²⁷ the *Book Buddies Early Literacy Screening* (BBELS),²⁸ and *The Howard Street Tutoring Manual*.²⁹ The validity of each word's grade-level placement was cross-checked for consistency within frequency bands in *The American Heritage Word Frequency Book*.³⁰ Words on the preprimer and primer word lists appear in at least three of the above sources. Words on the first-grade word list appear in at least two of the above sources and are unique to that specific grade level.

Preprimer, primer, and first-grade word lists were piloted in Spring 2001 with 427 students in 39 kindergarten classrooms in different schools and school divisions within the Commonwealth of Virginia. Two additional sets of word lists were piloted in Fall 2001 with 311 kindergarten students in 31 kindergarten classrooms. Furthermore, different forms of the PALS word lists have been piloted over the past five years with over 7,500 students in 246 first-, 194 second-, and 80 third-grade classrooms from over 55 different school divisions across all eight regions of the Commonwealth of Virginia. Student scores generated from all pilot studies were used to assess the reliability and validity of the word lists.

Analysis. Individual words and word lists were analyzed using the following criteria: (a) teacher feedback, (b) item means (level of difficulty), (c) item-to-total correlations, and (d) Cronbach's α . Words and/or word lists were considered for removal if they received negative feedback from more than two teachers in the pilot sample, were too easy or too difficult, had low item-to-total correlations, or had α coefficients lower than .80. Words with

low item-to-total correlations, little to no variance in response patterns, and/or negative feedback from teachers were substituted with words that had higher item-to-total correlations, moderate variance, and more positive teacher feedback. In a few isolated cases, plural endings were changed to singular. Currently, three sets of word lists with good evidence of reliability and validity are rotated across PALS screening windows.

Feedback from the Field

In addition to the formal feedback we solicit from reviewers, the PALS office continually seeks informal feedback from the field. During each spring screening window we post a survey on the PALS website (pals.virginia.edu) to solicit feedback from teachers in the field. For example, response rates to specific questions on the Spring 2001 survey ranged from 200 to 800 teachers who participated in the EIRI and who screened their students with either PALS-K or PALS 1–3. In Spring 2001, approximately 533 teachers rated PALS-K tasks on the ease of administration and scoring, the clarity of directions,

and the information gained from the screening. Open-ended comments were also invited.

The results from the survey and qualitative comments responding to an open-ended question were consistent with comments received through the toll-free phone line. That is, with regard to clarity of directions, ease of administration and scoring, and information gained from screening, most teachers (81% to 98% across all subtasks) rated PALS-K tasks good (4) to excellent (5) on a rating scale of one to five. In 2003, 2,011 teachers responded to a brief survey designed primarily to assess the usefulness of various PALS reports and Web site features. Between 71% and 80% of respondents rated class reports, class summary sheets, score history reports, and student summary reports as “very useful,” 2% or fewer of respondents rated any of these reports as “not useful.”

Outside Review

The Code of Fair Testing Practices in Education (1988) defines the obligations of professionals who

Table 4 VDOE Advisory Review Panel

Dr. Mary Abouzeid McGuffey/Tempo State Outreach Program University of Virginia	Lara Major Loudoun County Public Schools Ashburn, VA
Kate Bisset Fairfax County Public Schools Alexandria, VA	Alecia Morgan Preschool Education Alexandria, VA
Kelly Carper Fairfax County Public Schools Fairfax, VA	Mary Pace Aquinas School, Diocese of Arlington Woodbridge, VA
Dorothy Donat Augusta County Public Schools Fishersville, VA	Dr. Beth Roberts Mary Baldwin College Staunton, VA
Amy Harlan Danville County Public Schools Danville, VA	Carol Shields Pittsylvania County Public Schools Danville, VA
Mary Kate Long Charles City County Public Schools King George, VA	Charlotte Tucker Amherst County Public Schools Amherst, VA

Table 5 External Reviewers

Dr. Nicholas Bankson <i>Professor of Communication Sciences & Disorders</i> James Madison University Harrisonburg, Virginia
Dr. Susan Brady <i>Professor of Psychology</i> University of Rhode Island & Haskins Laboratories New Haven, Connecticut
Dr. Francine Johnston <i>Associate Professor of Reading</i> University of North Carolina-Greensboro
Dr. Frank Vellutino <i>Professor of Psychology & Director, Child Research & Study Center</i> State University of New York at Albany

undertake the process of creating an assessment instrument. Included among these obligations are procedures that minimize the potential for bias or stereotyping. The potential for bias can be minimized if assessment tools are carefully evaluated.³¹ Procedures that protect against inappropriate instrument content include the use of an advisory review panel and an external evaluation.

Advisory Review Panel

To evaluate the appropriateness of PALS' content, we sought the opinions of an advisory review panel, appointed by the Virginia Department of Education. The panel consisted of primary grade teachers, reading specialists, speech teachers, instructional coordinators, and educators from the Commonwealth of Virginia. Members of this panel are listed in Table 4, along with their affiliations at the time they reviewed PALS-K.

The advisory review panel evaluated the materials that are part of the PALS-K screening. Members completed review forms in which they appraised (a) the content of the assessment, (b) the content of the teacher's manual, (c) the directions for administration and scoring, (d) the content of the screening instrument, and (e) the graphic qualities of the mate-

rials. The review panel was further asked to suggest changes or deletions of items. Additionally, a measurement professional from outside the University of Virginia reviewed PALS-K materials to verify that the items and directions for administering and scoring met the minimum standards of assessment.

External Review

In addition to the opinions of the advisory review panel, the Virginia Department of Education sought the opinions of several external reviewers, all of whom are national experts in the fields of reading, communication sciences, or psychology. The first PALS technical manual and report,³² detailing the psychometric qualities of PALS and first-year results, as well as PALS materials and teacher's manuals, were sent to prominent researchers whose charge was to determine the technical soundness of PALS as a valid and reliable instrument for the EIRI. The opinions of these outside reviewers were presented to the Virginia Department of Education in March 1999. The judgments of these reviewers were favorable. Copies of the reviews can be obtained from the Virginia Department of Education. External reviewers are listed in Table 5. An additional, independent review of PALS can be found in *Early Reading Assessment*.³³

Section IV

Establishing Summed Score Criteria and Benchmarks

In the following sections, we describe the process through which PALS-K benchmarks were established.

Criteria and benchmarks for PALS-K were derived from several sources:

- nine years of research using similar tasks with struggling readers in a central Virginia early intervention program;
- statewide kindergarten and first-grade PALS data generated from the first eleven cohorts of Virginia's EIRI;
- data gathered from pilot samples between 2000 and 2004 with approximately 4,000 kindergarten students in the Commonwealth of Virginia;
- theoretical assumptions based on the reading research literature.

Benchmarks reflect raw scores for each PALS task, based on available data sources. The sum of these benchmark scores for the core variables equals the Summed Score criterion. These benchmarks and criteria are re-evaluated based on analyses of each year's statewide PALS-K data and results of ongoing pilot studies.

In November 2002 we conducted a formal standard-setting procedure to verify PALS benchmarks. Standard setting refers to the process used by instrument developers to help establish, or in this case to verify, benchmarks or levels of performance that reflect 'minimal competence.' In standard setting, expert judges evaluate each individual task or item and state whether they believe that the student who is minimally competent would respond correctly. We assembled panels of experts in reading from throughout the Commonwealth (one panel

of 20 judges was invited for each grade level, K through 3). Each panel of judges spent a full day in Charlottesville evaluating individual task items from all PALS materials.

We evaluated standard-setting judges' mean scores for PALS tasks against two sources of information: our current benchmarks, and statewide data from the most recent screening windows. In virtually all cases, standard-setting judges' scores were comparable to current benchmarks (i.e., within one standard deviation), and moreover fell at approximately the bottom quartile, which has traditionally been the approximate range of students identified by PALS-K. For these reasons, we decided that standard-setting judges' evaluations supported PALS benchmarks as appropriate.

Criteria and benchmarks for the 1997–99 PALS (the previous version of PALS, used with kindergarten and first-grade students only) were derived from more than six years of research using PALS measures on more than 750 at-risk children in a central Virginia early intervention program.³⁴ Since this research was limited to central Virginia, new benchmarks were established based on a statewide sample of 37,072 kindergarten children in the Commonwealth of Virginia in the fall of 1997. These benchmarks were confirmed through analyses conducted using scores from 50,949 kindergartners in Fall 1998; 53,256 kindergartners in Fall 1999; 74,054 kindergartners in Fall 2000; 50,127 kindergartners in Spring 2001; 65,036 kindergartners in Fall 2001; 74,928 kindergartners in Spring 2002; and 83,099 kindergartners in Fall 2003.

The core variables representing the emergent-literacy construct measured by PALS-K underwent principal component analysis (PCA) in order to obtain a single-factor score for each child. This was done for the ultimate purpose of calculating quartiles for that single index. The factor loadings from which the factor scores were created were relatively homogeneous, suggesting that each subtask similarly influenced the factor score. As a result, a different and simpler index was created. The scores obtained by each child on the core tasks were summed. To assess the validity of the Summed Scores, they were correlated with the more precise factor scores. That correlation coefficient was 0.99, indicating that the Summed Score captured nearly everything. Quartiles were then calculated for the Summed Scores for kindergarten using the statewide sample.

Benchmarks reflect milestones established in part by using means and standard deviations from students NOT in the bottom quartile. Benchmarks were determined initially by subtracting one standard deviation from the mean score for students ranking above the bottom quartile and by making further adjustments based on modal data for each task. Finally, we always evaluate benchmarks subjectively to make certain that decisions we have made empirically reflect sensible targets that are consistent with literacy acquisition theory.

Benchmarks and Discriminant Analysis (DA)

To verify PALS-K benchmarks statistically, we subject statewide data to discriminant analyses (DA). This allows us to assess the extent to which PALS variables reliably discriminate between groups of students who are or are not identified as needing additional services, based on their PALS-K Summed Score. The primary goal of DA is to isolate statistically the dimensions on which groups differ, based on a set of variables (i.e., PALS-K subtask scores).

Discriminant function analyses based on the subtasks included in the Summed Score correctly classified as Identified or Not-identified 96% of students in Fall 2009 and 98% of students in Spring 2010, based on their subtask scores. This suggests that the task scores used to create the Summed Score produce a discriminant function (a linear combination of these variables) that classifies students as Identified or Not-identified, using mathematical measures to isolate the dimensions that distinguish the groups. The abstract (or mathematical) classifications have consistently demonstrated a very high correspondence to PALS classification. Since the inception of PALS, similar DA analyses have consistently classified 93% to 98% of students correctly as Identified or Not-identified.

Section V

Technical Adequacy

In this chapter, we provide an overview of the students who have made up the PALS pilot and statewide samples, and then describe the technical adequacy of PALS-K in terms of validity and reliability.

Standards for test construction, evaluation, and documentation, as outlined in the Standards for Educational and Psychological Testing (1999), prepared by the American Educational Research Association (AERA), American Psychological Association (APA), and National Council on Measurement in Education (NCME), were carefully followed throughout the development of PALS-K. We made special efforts to satisfy all the major criteria for acquiring and reporting technical data. In addition, we have attended carefully to the assessment criteria spelled out in various policy initiatives (e.g., the Reading First requirement of the No Child Left Behind Act, Race to the Top). Specifically, Reading First guidelines suggest that assessment tools should serve four assessment purposes: (a) screening, (b) diagnosis, (c) progress monitoring, and (d) outcome evaluation. Moreover, states are encouraged to use assessments that target five core reading areas: (a) phonemic awareness, (b) phonics, (c) fluency, (d) vocabulary, and (e) comprehension.

In general, PALS provides an assessment tool that clearly meets screening and diagnostic assessment purposes, with the mid-year form providing some degree of progress monitoring. Originally designed as a screening tool for identifying children who were behind in the acquisition of important literacy fundamentals, PALS was not intended to serve as a tool to assess outcomes. The screening and diagnostic aims of PALS-K are readily apparent in the nature of information teachers glean from its subtasks. The focus of PALS-K on the five core reading areas is also evident

in the direct and instructionally relevant assessment of these literacy fundamentals (see the conceptual framework of PALS-K in Table 3) in the subtasks of Rhyme Awareness, Beginning Sound Awareness, Alphabet Recognition, Letter Sounds, Spelling, Concept of Word, and Word Recognition in Isolation.

In subsequent sections, we describe the technical adequacy of PALS-K in terms of reliability and validity. First, we provide an overview of the students who have participated in PALS-K pilot and statewide samples.

Broad Representation of Students

Tasks, items, and benchmarks used in PALS-K are derived from analyses of PALS scores from more than 600,000 kindergarten students in schools participating in Virginia's EIRI between 1997 and 2006. The first ten cohorts of the EIRI provide ten statewide samples representing a diverse population.³⁵ Table 6 lists the total number of students screened with PALS-K in the sixteenth cohort of Virginia's EIRI (Fall 2012) by gender, socioeconomic status (SES), and race/ethnicity.

Pilots

Data on the development, refinement, and technical adequacy of PALS-K items and scoring procedures were obtained from statewide data collected on kindergarten children since Fall 1997 and from large-scale pilots conducted in Spring 2001, Fall 2001, Spring 2002, Spring 2003, and Spring 2004. PALS-K statewide samples have ranged from 37,072 to 87,451. Table 7 contains a summary of the participants in pilots.

In pilots, we made efforts to ensure that samples approximated statewide school enrollments in terms of gender, race/ethnicity, and SES. Table 8 summarizes the demographics of Spring 2003 and Spring 2004 pilot samples. For each demographic category, the percentage in the total pilot sample is compared to the percentage in the total statewide kindergarten enrollment. With the possible exception that our pilot sample included fewer students from low-poverty areas and more students from high-poverty areas, the pilot samples generally mirrored the demographics of statewide enrollment.

Summary Statistics

Currently, students screened with PALS-K are identified as in need of additional instruction based on their Summed Score, which is the sum of six subtask scores: Rhyme Awareness, Beginning Sound Awareness, Alphabet Recognition, Letter Sounds, Spelling, and Concept of Word Post-test Word List. Tables 9 and 10 summarize descriptive data for the Summed Score for statewide samples. Spring scores are not included here because only students identified in the fall were required to be rescreened in the spring until Spring

**Table 6 Demographics of Virginia's 17th EIRI Cohort: Kindergarten, Fall 2013
(*n* = 80,955)**

Demographic Category		<i>n</i> (%)
GENDER	Males	41,499 (51.2%)
	Females	39,456 (48.7%)
ECONOMIC STATUS	Eligible for FRPL	37,200 (46.0%)
	Not eligible for FRPL	37,751 (46.6%)
	Unknown economic status	6,004 (7.4%)
RACE/ETHNICITY	Black or African American	21,063 (26.0%)
	White (Not of Hispanic Origin)	42,264 (52.2%)
	Hispanic	10,772 (13.3%)
	Asian	3,116 (3.9%)
	American Indian/Alaska Native	472 (0.6%)
	Native Hawaiian/Other Pacific Islander	291 (0.4%)
	Two or more races	2,977 (3.7%)

FRPL = Free or reduced price lunch.

Table 7 Pilot Participation Summary: Spring 2001 through Mid-Year 2005

Pilot	# Schools	# Teachers	# Students
Spring 2001	136	156	1,772
Fall 2001	108	142	2,205
Spring 2002	62	163	2,849
Spring 2003	71	72	1,565
Fall 2003	63	173	2,789
Spring 2004	71	252	3,924
Mid-Year 2005	12	44	193

Table 8 Pilot Sample Demographics Compared to Statewide Enrollment: Spring 2003 and Spring 2004

Demographic Category		Spring 2003 Pilot	2002–03 Statewide Enrollment	Spring 2004 Pilot	2003–04 Statewide Enrollment
GENDER	Males	50.8%	51.3%	51.2%	51.5%
	Females	49.2%	48.7%	48.8%	48.5%
SES*	Low FRPL	33.1%	29.1%	16.6%	30.8%
	Med-Low FRPL	17.3%	25.0%	28.6%	25.6%
	Med-High FRPL	32.1%	22.2%	27.5%	22.5%
	High FRPL	17.5%	23.7%	27.3%	20.5%
ETHNICITY	Black	21.6%	27.0%	26.3%	26.9%
	White	66.7%	61.3%	65.9%	60.4%
	Hispanic	5.2%	6.1%	4.1%	6.5%
	Asian/Pacific Islander	2.9%	4.5%	2.1%	4.6%
	American Indian/ Alaska Native	0.9%	0.3%	0.4%	0.5%

*SES is expressed in terms of school-level free and reduced price lunch (FRPL) counts.

2002; thus spring samples prior to 2002 are smaller. As displayed in Table 9, PALS-K identification rates have trended downward since 2003 to about 12% of kindergarten screeners identified as needing additional instruction in Fall 2011. In Table 10, the discrepancy between means and standard deviations for Identified and Not-identified groups highlights the clear distinction between these groups.

We examine and summarize PALS-K scores each year for indices of central tendency, internal consistency, and item reliability. We also conduct factor analyses

and discriminant function analyses to assess the validity of PALS-K tasks. The following sections contain a brief description of the technical adequacy of PALS-K in terms of reliability (the consistency of scores) and validity (the extent to which PALS-K is supported as a true measure of the construct of emergent reading).

Reliability

Reliability coefficients provide information about the consistency of test scores. Reliability may be assessed

Table 9 Number and Percentage of Students Identified Based on Summed Score

Date	Screened	Identified
Fall 1997	37,072	7,398 (20%)
Fall 1998	50,949	13,143 (26%)
Fall 1999	53,256	13,045 (24%)
Fall 2000	74,054	16,127 (22%)
Fall 2001	65,036	12,302 (19%)
Fall 2002	74,666	15,194 (20%)
Fall 2003	83,099	17,792 (21%)
Fall 2004	83,934	16,551 (19%)
Fall 2005	85,487	15,237 (18%)
Fall 2006	87,273	15,009 (17%)
Fall 2007	76,444	12,581 (16%)
Fall 2008	72,708	10,417 (14%)
Fall 2009	74,563	10,538 (14%)
Fall 2010	79,044	10,642 (13%)
Fall 2011	80,138	9,920 (12%)
Fall 2012	81,513	10,517 (13%)
Fall 2013	80,955	10,088 (12%)

Table 10 Means and Standard Deviations for Summed Scores for Kindergarten Students Identified and Not-identified

Date	Summed Score Mean (sd)	
	ID	Not-ID
Fall 1998	16.93 (6.65)	55.24 (16.34)
Fall 1999	16.97 (6.72)	55.52 (16.47)
Fall 2000	17.07 (6.80)	56.78 (16.58)
Fall 2001	18.85 (7.21)	67.62 (23.34)
Fall 2002	17.27 (6.80)	58.29 (16.99)
Fall 2003	16.82 (6.96)	60.50 (19.39)
Fall 2004	17.08 (6.89)	61.09 (19.12)
Fall 2005	17.00 (6.90)	63.31 (19.49)
Fall 2006	16.76 (7.04)	64.07 (19.36)
Fall 2007	17.08 (6.79)	64.82 (19.53)
Fall 2008	16.99 (6.91)	65.95 (19.27)
Fall 2009	17.22 (6.85)	66.82 (19.61)
Fall 2010	17.11 (6.86)	67.06 (19.48)
Fall 2011	17.36 (6.71)	68.13 (19.57)
Fall 2012	17.00 (7.01)	67.50 (19.40)
Fall 2013	17.40 (6.69)	67.88 (19.41)

Data are not provided here for the initial PALS cohort in 1997, as identification was based not on a Summed Score criterion but on passing a certain number of subtasks within each domain. Scores in Fall 2001 were higher because Concept of Word was included in Summed Score.

by comparing the scores of individuals taking the same test on different occasions (test-retest reliability), taking equivalent forms of the test (equivalent forms reliability), or, when it is not practical to assess individuals on two separate occasions, to examine the internal consistency of the scale (e.g., split-half reliability). Reliability evaluates the error of measurement or the “true score” variance. We assess three aspects of PALS’ reliability: test-retest reliability, internal consistency (subtask reliability), and the consistency and accuracy of scoring (inter-rater reliability). Internal consistency was assessed using Cronbach’s alpha;³⁶ these results are reported in the following sections. Inter-rater reliability was assessed by having tasks scored and tabulated by multiple raters.

Test-retest Reliability

Test-retest reliability was assessed in Fall 2002 with a sample of 473 students. In this study, teachers administered PALS-K a second time to a randomly selected sample of their students. These students were tested again at least one week, but no more than two weeks, after their initial screening. We then computed Pearson correlations between scores on the two administrations as an indicator of test-retest reliability. Test-retest reliabilities, which ranged from .78 to .95, are presented in Table 11.

Subtask Reliability

Reliabilities for PALS subtasks were determined for gender, SES, race/ethnicity, and region using data generated from statewide samples from 1998 to 2007. Task reliabilities were determined using Cronbach’s

alpha.³⁷ Table 12 displays the alpha coefficients for PALS-K tasks by gender, SES, and race/ethnicity, based on statewide samples from fall screenings from 1998 through 2007. We do not report spring screening results here for years prior to 2002, as during those years only students identified in the fall were required to be rescreened in the spring. Thus, spring samples prior to Spring 2002 were not representative of statewide samples overall.

Reliabilities for PALS subtasks by gender and race/ethnicity are presented in Table 13 for Fall 2007 and later. These reflect the new categories used by the Virginia Department of Education to describe children’s ethnicity. We also began consistently disaggregating SES to a finer degree starting Fall 2007, using deciles (10 equal groups) instead of quartiles (4 equal groups). For Fall 2008 and Spring 2009, Cronbach’s alpha averaged .86 (range = .78 to .88) across the ten decile groups of school-level SES (based on free or reduced lunch counts).

Inter-rater Reliability

Inter-rater reliability coefficients provide evidence that different individuals score a particular task the same way. To determine the inter-rater reliability of PALS-K, scores for various PALS-K tasks from two different raters (or scorers) were compared. The most extensive assessments of inter-rater reliability were conducted in Fall 1997 and Spring 1999. In these studies, one person administered the PALS-K subtasks while a second person observed and scored the tasks simultaneously but independently. Each person

Table 11 Test-retest Reliability Expressed as Pearson Correlations, Fall 2002	
PALS-K Task	Test-retest Correlation (<i>n</i>)
Group Rhyme Awareness	.81 (<i>n</i> = 467)
Group Beginning Sound Awareness	.78 (<i>n</i> = 470)
Alphabet Recognition	.92 (<i>n</i> = 472)
Letter Sounds	.88 (<i>n</i> = 473)
Spelling	.89 (<i>n</i> = 473)
Concept of Word (total)	.92 (<i>n</i> = 473)
Summed Score	.95 (<i>n</i> = 472)

Table 12 Task Reliability across Demographic Categories: Kindergarten, Entire Sample

	Entire Sample	Female	Male	SES 1	SES 2	SES 3	SES 4
Fall 1998 (<i>n</i> = 50,949)	.83	.83	.83	.85	.85	.83	.81
Fall 1999 (<i>n</i> = 53,256)	.83	.83	.83	.85	.85	.83	.81
Fall 2000 (<i>n</i> = 74,054)	.84	.84	.84	.83	.84	.84	.83
Fall 2001 (<i>n</i> = 65,036)	.87	.87	.87	.86	.87	.87	.86
Spring 2002 (<i>n</i> = 66,658)	.82	.81	.83	.83	.82	.81	.80
Fall 2002 (<i>n</i> = 73,427)	.84	.84	.84	.84	.84	.84	.84
Spring 2003 (<i>n</i> = 77,539)	.87	.86	.87	.87	.87	.86	.85
Fall 2003 (<i>n</i> = 83,099)	.89	.89	.89	.86	.86	.85	.84
Spring 2004 (<i>n</i> = 84,311)	.87	.86	.87	.87	.87	.86	.85
Fall 2004 (<i>n</i> = 83,934)	.85	.85	.85	.86	.85	.85	.84
Spring 2005 (<i>n</i> = 84,830)	.86	.85	.86	.87	.86	.86	.84
Fall 2005 (<i>n</i> = 85,723)	.86	.86	.86	.84	.85	.85	.86
Spring 2006 (<i>n</i> = 87,510)	.87	.86	.87	.87	.87	.87	.85
Fall 2006 (<i>n</i> = 87,273)	.86	.86	.86	.85	.85	.86	.86
Spring 2007 (<i>n</i> = 88,376)	.86	.85	.86	.87	.86	.85	.84

SES based on quartiles of free/reduced lunch (F/R) at the school level. For the 2006–07 school year, these quartiles were: SES 1 = 57.7% F/R or greater; SES 2 = 39.4–57.7% F/R; SES 3 = 19.2–39.4% F/R; SES 4 = 0–19.2% F/R.

	African American	Asian & Pacific Islander	Caucasian	Hispanic	Native American	Other
Fall 1998	.80	.83	.84	.82	.80	.85
Fall 1999	.80	.83	.84	.82	.80	.85
Fall 2000	.83	.84	.84	.84	.83	.83
Fall 2001	.86	.88*	.88	.88	.86	.88
Spring 2002	.83	.83	.80	.86	.80	.82
Fall 2002	.83	.84	.84	.84	.84	.85
Spring 2003	.87	.87	.85	.89	.84	.86
Fall 2003	.84	.86	.86	na	.86	.86
Spring 2004	.87	.86	.85	.88	.86	.87
Fall 2004	.84	.86	.85	.84	.86	.86
Spring 2005	.86	.85	.84	.88	.86	.85
Fall 2005	.85	.87	.86	.85	.85	.86
Spring 2006	.87	.85	.85	.88	.81	.87
Fall 2006	.85	.87	.86	.85	.85	.86
Spring 2007	.86	.85	.84	.87	.84	.85

*Asian & Pacific Islander group was counted separately as “Asian” or “Hawaiian or Other Pacific Islander” in Fall 2001; Cronbach’s alpha was .88 for both groups.

Table 13 Task Reliability across Demographic Categories: 2007–12

	Entire Sample	Male	Female
Fall 2007 (<i>n</i> = 76,444)	.86	.86	.86
Spring 2008 (<i>n</i> = 76,954)	.87	.87	.86
Fall 2008 (<i>n</i> = 72,708)	.88	.88	.88
Spring 2009 (<i>n</i> = 76,735)	.83	.84	.81
Fall 2009 (<i>n</i> = 74,563)	.88	.88	.88
Spring 2010 (<i>n</i> = 79,220)	.84	.85	.83
Fall 2010 (<i>n</i> = 79,044)	.88	.88	.88
Spring 2011 (<i>n</i> = 76,444)	.83	.84	.82
Fall 2011 (<i>n</i> = 80,138)	.86	.86	.86
Spring 2012 (<i>n</i> = 80,241)	.84	.85	.83
Fall 2012 (<i>n</i> = 81,519)	.87	.87	.86
Spring 2013 (<i>n</i> = 82,004)	.85	.86	.84
Fall 2013 (<i>n</i> = 80,955)	.87	.86	.87
Spring 2014 (<i>n</i> = 81,069)	.86	.86	.85

	Black	White	Hispanic	American Indian/ Alaska Native	Asian	Native Hawaiian/Other Pacific Islander	Ethnicity Unspecified
Fall 2008	.87	.88	.88	.88	.88	.88	.88
Spring 2009	.84	.81	.84	.81	.82	.85	.82
Fall 2009	.88	.88	.88	.89	.89	.89	.88
Spring 2010	.85	.83	.86	.85	.82	.83	.83
Fall 2010	.88	.88	.88	.87	.88	.89	.88
Spring 2011	.83	.82	.84	.83	.83	.86	.89
Fall 2011	.86	.86	.85	.86	.87	.86	.89
Spring 2012	.84	.84	.85	.85	.84	.85	—
Fall 2012	.86	.86	.86	.88	.87	.88	—
Spring 2013	.85	.84	.86	.86	.83	.82	—
Fall 2013	.86	.86	.86	.87	.87	.87	—
Spring 2014	.86	.85	.87	.86	.86	.86	—

administering or scoring tasks in these studies experienced the same training provided to all teachers using PALS: they read the PALS teacher's manual and viewed the PALS training video prior to administration. As shown in Table 14, inter-rater reliability coefficients were consistently high (range: .96–.99), suggesting that PALS-K can indeed be administered and scored reliably.

Internet Data Entry Reliability

In the Commonwealth of Virginia, teachers enter PALS scores via the Internet into a password-protected, securely encrypted database. The reliability of score entry into the Internet database is checked regularly against a randomly selected sample of the original hand-scored Class Summary sheets. In Spring 2003, we compared a 10% sample of these Class Summary Sheets against the PALS database. Based on a sample of 5,931 students' score entries, which consisted of 74,612 individual data points, we found 708 errors, reflecting an overall accuracy of Internet data entry of 99.1%.

Validity

In general terms, validity refers to the extent to which one can trust that a test measures what it is intended to measure. More specifically, a test must be assessed for evidence of validity in relation to the specific purpose for which it is used in a given population. Thus, for PALS-K, three types of validity have been assessed through our pilot studies and our examination of statewide PALS data over the past five years. In the following sections, we describe evidence of PALS' (a) content validity, (b) criterion-related validity, both predictive and concurrent, and (c) construct validity. Finally, to provide further evidence of validity, we assessed the differential item functioning of PALS tasks for different groups of students. A report of those results follows as well.

Content Validity

Content validity is the degree to which the sample of items and tasks provides a relevant and represen-

Table 14 Inter-rater Reliabilities Expressed as Pearson Correlation Coefficients for PALS Tasks

PALS-K Task	Date	Correlation (<i>n</i>)
Rhyme Awareness	Fall 1997	K & 1: .99 (<i>n</i> = 134)
	Spring 1999	K & 1: .96 (<i>n</i> = 154)
Beginning Sound Awareness	Fall 1997	K & 1: .99 (<i>n</i> = 122)
	Spring 1999	K & 1: .99 (<i>n</i> = 154)
Alphabet Recognition	Fall 1997	K & 1: .99 (<i>n</i> = 122)
	Spring 1999	K & 1: .99 (<i>n</i> = 154)
Letter Sounds	Fall 1997	K & 1: .99 (<i>n</i> = 121)
	Spring 1999	K & 1: .98 (<i>n</i> = 154)
Spelling	Fall 1997	K & 1: .99 (<i>n</i> = 130)
	Spring 1999	K & 1: .99 (<i>n</i> = 154)
Concept of Word (total score)	Fall 2001	K: .97 (<i>n</i> = 110)
Word Recognition in Isolation	Fall 2000	*Preprimer: .99 (<i>n</i> = 51) *Primer: .99 (<i>n</i> = 52) *First Grade: .98 (<i>n</i> = 45)

p < .01 for all correlations, * indicates level of word list; inter-rater reliability for word lists was assessed in Fall 2000 using students in first through third grades.

tative sample of the content addressed (Gronlund, 1985). To ensure that the phonological awareness tasks in PALS-K have ample content validity, special care was taken to select items that represent the subject matter being assessed. The research literature identified two specific levels of phonological awareness: rhyme awareness and phonemic awareness.³⁸ Both of these levels are represented in the Rhyme Awareness and Beginning Sound Awareness tasks of PALS-K.

To ensure that the literacy tasks would have ample content validity, special care was taken to select items that represent the literacy subject matter being assessed. To assess alphabet recognition, for example, all 26 letters of the alphabet were included. To assess knowledge of letter sounds, all letters were included except Q and X, letters that are difficult to pronounce in isolation. To assess word recognition, we included representative word lists from the beginning, middle, and end of first grade. Finally, to assess concept of word, contextual finger-point reading with specific

line-by-line behavioral criteria was included. A more detailed explanation of the content validity of PALS-K tasks can be found in this technical reference under Item Development.

Criterion-related Validity

Criterion-related validity determines whether assessment scores are related to one or more outcome criteria.³⁹ There are two types of criterion-related validity: predictive, in which an assessment is used to predict future performance; and concurrent, in which assessment results are compared to performance on a different assessment administered at approximately the same time. Both forms of validity have been assessed for PALS-K.

Predictive Validity. The predictive validity of PALS-K has been assessed in two ways. First, PALS scores from the fall were compared with Stanford Achievement Test⁴⁰ scores obtained during the spring of the same school year. When PALS was developed, the administration of the Stanford-9 was required in the Commonwealth of Virginia in grades 3, 5, 8, and 11. In addition, the kindergarten and first-grade versions of the Stanford-9 contain three subtests that are similar, though not identical, to several PALS tasks: Sounds and Letters, Word Reading, and Sentence Reading. In Fall 1998, 74 kindergartners from one school division were screened with PALS. None of the students were provided additional instruction apart from that which all students receive during the school year. The same 74 students were given the Stanford-9 at the end of the school year, in Spring 1999. Fall PALS Summed Scores and all PALS subtask scores were significantly correlated with spring Stanford-9 scaled scores ($p < .001$). The correlation between fall PALS Summed Scores and spring Stanford-9 Total Reading scaled scores was .70.

Significant amounts of variance in the kindergarten Stanford-9 Total Reading scaled scores were explained by the five core PALS subtasks collectively (Rhyme Awareness, Beginning Sound Awareness, Alphabet Recognition, Letter Sounds, and Spelling). In regression equations, the proportion of variance explained

by the total model was 50% ($p < .001$), and the adjusted R^2 was .47. PALS fall Summed Scores also predicted spring Stanford-9 scaled scores for all three Stanford-9 subtests: Sentence Reading, Word Reading, and Letters and Sounds ($p < .001$). The adjusted R^2 for Stanford-9 Word Reading was .54 ($p < .001$).

A second assessment of predictive validity involves an examination of the relationship between PALS-K scores from a current administration and future PALS scores. For example, we found significant ($p < .001$ in all cases) and medium to medium-high correlations between kindergarten students' Summed Scores from Fall 2000 and later PALS-K scores from spring of their kindergarten year ($r = .56$), as well as with PALS 1–3 Entry Level scores from the fall ($r = .67$) and spring ($r = .53$) of their first grade year. The shared variance evident in these correlations offers some evidence of the predictive power of PALS-K Summed Scores relative to future PALS scores.

We also examined the predictive power of individual subtask scores from PALS-K in Fall 2000 relative to future PALS-K and PALS 1–3 scores. Regression equations using the subtask scores making up the PALS-K Summed Score (Rhyme Awareness, Beginning Sound Awareness, Alphabet Recognition, Letter Sounds, and Spelling) as independent variables yielded adjusted R^2 values of .33 in predicting Spring 2001 Summed Scores, .45 in predicting Fall 2001 PALS 1–3 Entry Level Scores, and .30 in predicting Spring 2002 PALS 1–3 Entry Level Scores. Thus the amount of variation in future PALS scores that could be predicted based on their relationship to PALS-K subtask scores ranged from 30% to 45% for the subsequent spring and the following fall and spring of first grade.

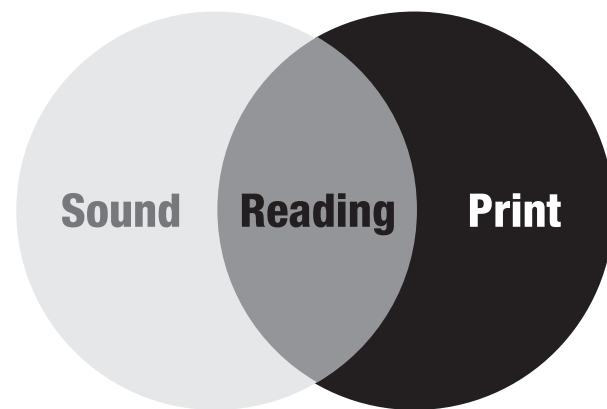
In a study of the predictive validity of PALS-K ($n = 61,124$), discriminant analysis was used to assess the relationship between the 2012 Reading SOL scores in the spring of third grade and the students' spring PALS-K scores three years earlier. The discriminant analysis correctly classified 85% of students according to their pass-fail status on the SOL.

The Reading SOL scores and the spring PALS-K summed scores were moderately correlated ($r = .43$). The individual tasks that had the highest correlation with the SOL scores were Concept of Word ($r = .43$), Spelling ($r = .35$), and letter sound knowledge ($r = .33$). Although approximately 77 thousand kindergarteners had been assessed using PALS-K in the spring of 2009, only 80% of the students had 2012 Reading SOL scores three years later. The students that remained in the sample had higher spring PALS-K scores ($M_s = 94.2$ vs. 82.8 , $d = 0.97$) and were more likely to be White (58% vs 50%) compared to the kindergarten students who had taken PALS-K and were not matched in the sample. With the limitations noted above, these correlations provide some evidence that PALS-K scores may account for a portion of the variation in a child's SOL scores three years later.

Concurrent Validity. Concurrent validity is the extent to which outcomes obtained from a particular measure are consistent with some independent standard.⁴¹ The independent standard against which PALS was compared was the Stanford-9 (1997). Again the three Stanford-9 subtests that are similar to PALS tasks (Sounds and Letters, Word Reading, and Sentence Reading) were administered in Spring 1999 to 137 kindergartners, who had also been given PALS two weeks earlier.

The correlation between the end-of-year kindergarten PALS Summed Score and the Total Reading scaled score of the Stanford-9 was medium to high and significant ($r = .72$, $p < .001$). The correlations between the PALS Summed Score and the three Stanford-9 subtest scaled scores were also medium to high and significant (Sounds and Letters, $r = .79$; Word Reading, $r = .74$; and Sentence Reading, $r = .58$). Correlations between the PALS Summed Score and the Stanford-9 raw scores were similar: medium to high and significant (Total Reading, $r = .79$; Sounds and Letters, $r = .80$; Word Reading, $r = .78$; Sentence Reading, $r = .56$). Consistently medium to high correlations provide evidence of the concurrent validity of PALS with the Stanford-9 when administered at the end of kindergarten.

Figure 1 PALS Theoretical Model



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.⁴²

The theoretical model on which PALS was based is illustrated in Figure 1. As demonstrated there, PALS was designed to assess children's knowledge of sound and print, and includes tasks that assess the wedding of the two. The pronunciation of letter sounds, the ability to invent a phonetically plausible spelling, and the ability to match speech to print and to recognize words out of context all require the application of both sound and print knowledge.

We tested the theoretical model illustrated in Figure 1 several ways. First, we conducted principal components analyses (PCA) on PALS data to verify the underlying factor structure. Second, we conducted discriminant analyses (DA) on PALS data to determine the extent to which group membership (i.e., Identified versus Not-identified as needing additional services) could be predicted accurately from PALS subtask scores. Third, we conducted receiver-operating characteristic (ROC) curve analysis to evaluate the diagnostic accuracy of PALS-K.

Principal Components Analysis (PCA). We tested the theoretical model illustrated in Figure 1 by subjecting the first-year PALS results to principal

components analysis (PCA). In the first year, PCA for the entire sample yielded one factor with an eigenvalue of 5.20. This factor represented the inter-relationship between sound and print. The same unitary factor was also found using kindergarten data only (eigenvalue of 4.92) and first-grade data only (eigenvalue of 4.05). The one-factor solution suggested that PALS measures a unitary trait: emergent literacy. These results are in keeping with Perney et al.'s (1997) research that also yielded a single factor. In Fall 1997, the single PALS factor accounted for 58% to 74% of the total variance in the children's scores on all the tasks in both the phonological awareness and literacy components of PALS for the entire sample, for kindergarten, and for first grade.⁴³

This unitary factor was replicated using Fall 1998 and Fall 1999 PALS results. Principal components analysis again yielded a single eigenvalue greater than one for the entire sample, for kindergarten, and for first grade. Factor loadings from the second and third year were similar to the first: five core variables (Rhyme Awareness, Beginning Sound Awareness, Alphabet Recognition, Letter Sounds, and Spelling) defined the construct. Factor loadings for Letter Sounds and Spelling were consistently large and account for most of the construct. Factor loadings for Rhyme Awareness were the smallest. This pattern stayed the same for the entire sample, for kindergarten only, and for first grade only.

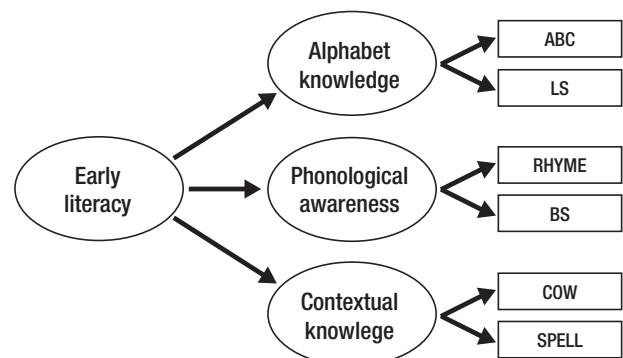
The PALS theoretical model was tested again using 2000–01 statewide data ($n = 74,054$) and subsequently using 2001–02 statewide data ($n = 65,036$), to see whether the newly configured Concept of Word task would also load onto one factor. Principal components analysis yielded one factor with an eigenvalue greater than one for each year. Even with Concept of Word included in the analysis, factor loadings are similar to previous factor loadings: all PALS-K tasks, including Concept of Word, define the construct. Factor loadings for Letter Sounds and Spelling remain the largest and account for most of the construct, while the factor loadings for Rhyme Awareness remain the smallest. This pattern stays

the same for fall and spring. The eigenvalue for the single factor resulting from PCA on PALS-K Spring 2011 data was 3.80. This factor accounted for nearly two-thirds (65%) of the total variance in the data set, a pattern that has remained consistent across several years of statewide data.

Factor Analysis. Exploratory and confirmatory factor analysis (CFA) were used to assess the factor structure of PALS-K with a sample of 2,844 first-time public-school kindergarteners.⁴⁴ The sample was randomly split in two to form exploratory and confirmatory samples. In a comparison of three models (e.g., a one factor model, a two-correlated factor model, and a hierarchical model), CFA results indicated that the data could best be represented by a second-order factor model. An overall general factor of early literacy influenced three first-order factors of alphabet knowledge, phonological awareness, and contextual knowledge. The factor model (see Figure 2) was found to replicate in the confirmatory hold-out sample providing further evidence of the model's generalizability.

In addition to assessing the factor structure of PALS-K, metric invariance of the instrument was investigated. Metric invariance is concerned primarily whether the estimated factor structure is statistically indistinguishable between defined groups

Figure 2 Second-order Factor structure of PALS-K



ABC = alphabet recognition. LS = letter sound knowledge. RHYME = rhyme awareness. BS = beginning sounds awareness. COW = concept of word. SPELL = spelling. Note. Error terms not shown.

of students. In this particular study, the factor model was tested comparing Spanish-speaking students who were English language learners (ELLs) with non-ELL students. Research supported the metric invariance of PALS-K between both groups of students lending support for the use of PALS-K with these populations.

Discriminant analyses (DA). The purpose of discriminant analysis is to determine whether test scores can discriminate accurately between groups of subjects if their group identity is removed. Because PALS is designed as a screening tool to identify students in need of additional reading instruction, we test this hypothesis each year by determining the extent to which a combination of PALS subtest scores accurately predicts membership in Identified and Not-identified groups.

In Spring 2010, discriminant analyses using the six PALS-K subtask scores that made up the Summed Score (Rhyme Awareness, Beginning Sound Awareness, Alphabet Recognition, Letter Sounds, Spelling, and Concept of Word-word list) yielded a function that was statistically significant in differentiating groups (as indicated by a statistically significant Wilks's lambda for the discriminant function). The same function accurately classified 98% of students as Identified or Not-identified. This classification rate has remained consistent since 1997, with discriminant analyses accurately classifying 93% to 98% of students. Table 15 summarizes DA results across the last thirteen PALS cohorts.

Receiver-operating characteristic (ROC) curve analysis. Receiver-operating characteristic (ROC) curve analysis is a tool for evaluating the performance of a diagnostic test in classifying subjects into one of two categories (e.g., at risk for future reading difficulties, not at risk for future reading difficulties) often used in laboratory testing, epidemiology, and psychology.⁴⁵ To estimate the classification accuracy using ROC curve analysis, an external indicator is used to evaluate how well PALS discriminates between those students who were identified as

needing additional intervention and those not identified as needing additional intervention.

As an external measure or risk status, students who scored at or below the 20th percentile of the Stanford Reading First⁴⁶ (SRF) were identified to be at risk for future difficulties. The SRF, an edition of the Stanford-10, is a national norm-referenced assessment. The SESAT/2 (Stanford Early School Achievement Test) multiple-choice reading subtest of the SRF, which was appropriate for students in the second half of kindergarten, was administered to the students by their teachers. The SRF focused on the five components of scientifically-based reading research: phonemic awareness, phonics, vocabulary development, reading fluency, and comprehension.

A total of 3,506 kindergarteners (males = 50.2%) in 73 schools in Virginia were assessed using both the PALS-K and the SRF in the spring of 2005. Of the sample, 45% were White, 48% were Black, 4% were Hispanic, and 3% were of another race/ethnicity. A majority of students (58%) were eligible for free or reduced-price lunch and 98% had no identified disabilities.

The Area Under the Curve (AUC) statistic of a ROC curve is an overall indication of diagnostic accuracy. AUC values of 1.00 indicate perfect classification accuracy whereas values of 0.50 indicate that the screener is no better than chance. The National Center on Response to Intervention uses an AUC value greater than .85, along with several other criteria, to indicate if a screener has convincing evidence for classification accuracy.⁴⁷ In our sample, PALS-K had an overall AUC of .91. In terms of disaggregated classification accuracy for White, Black, and Hispanic populations, PALS-K had AUC values of .91, .90, and .92, respectively. Findings indicate that PALS-K has excellent classification capabilities.

Together, the results of our PCA, CFA, DA, and ROC curve analyses continue to suggest that PALS-K assesses a single general construct associated with emergent reading and, further, that the combination of variables making up the PALS-K subtasks discriminates reliably

Table 15 Discriminant Analysis Results for Summed Score Tasks and Identification Status

Date	Wilk's lambda*	Students Classified Accurately
Fall 1997	.33	94%
Fall 1998	.34	96%
Fall 1999	.34	96%
Fall 2000	.21	95%
Fall 2001	.24	94%
Spring 2002	.33	96%
Fall 2002	.36	93%
Spring 2003	.31	97%
Fall 2003	.33	95%
Spring 2004	.32	97%
Fall 2004	.34	95%
Spring 2005	.34	96%
Fall 2005	.34	94%
Spring 2006	.33	98%
Fall 2006	.34	96%
Spring 2007	.34	97%
Fall 2007	.34	96%
Spring 2008	.34	98%
Fall 2008	.35	96%
Spring 2009	.35	97%
Fall 2009	.36	96%
Spring 2010	.34	98%
Fall 2010	.35	96%
Spring 2011	.36	98%

* $p < .001$ in all cases

between groups of students who are or are not identified as needing additional reading instruction.

Intercorrelations among PALS-K Tasks. A third source of evidence for a test's construct validity may be found in the intercorrelations among its subtests. We examined the intercorrelations among PALS-K task scores to assess the relationships among PALS-K tasks and, further, to verify that the pattern of intercorrelations is consistent across student subgroups (e.g., SES levels or race/ethnicity categories).

High correlations (above .80) are consistently obtained between the PALS-K Summed Score and Alphabet Recognition, Letter Sounds, and Spelling

task scores. In Fall 2011, for example, these correlations were .87 (Alphabet Recognition); .94 (Letter Sounds); and .88 (Spelling).

Medium-high correlations (between .60 and .79) are consistently obtained between Summed Scores and Beginning Sound Awareness ($r = .76$); Rhyme Awareness ($r = .60$); preprimer Word Recognition in Isolation ($r = .75$); and primer Word Recognition in Isolation ($r = .61$). In Fall 2011, Concept of Word-word list also correlated in the medium-high range ($r = .75$) with the Summed Score. This pattern of intercorrelations among PALS tasks administered to kindergartners in the fall has been consistent over the past years.

Patterns of intercorrelation among PALS-K tasks are also examined within subgroups of the statewide sample based on geographic region, gender, race/ethnicity, and SES. For example, the pattern of intercorrelation across regions of the state has consistently mirrored that of the entire sample. That is, for the last three years, all intercorrelations that were high in the entire sample were also high within each of Virginia's eight regions. The same can be said for intercorrelations that were medium-high and medium as well. The patterns of intercorrelation are also similar for males and females, racial/ethnic groups, and all levels of SES, suggesting that tasks in PALS-K behave in a similar manner for all students regardless of geographic region, gender, SES, or race/ethnicity.

The one exception to the pattern of consistency across groups emerged when we examined intercorrelations within groups of Identified versus Not-identified students. Generally, intercorrelations were lower within the Identified group than the Not-identified group; this is likely due to the restriction in the range of scores that naturally occurs when the Identified group is isolated.

Differential Item Functioning

Differential item functioning refers to the consistency of response to specific items or tasks across groups. The Mantel-Haenszel (MH) statistic can be

defined as the average factor by which the odds that members of one group will answer a question correctly exceed the corresponding odds for comparable members of another group. The MH statistic is a form of odds ratio.⁴⁸

To explore the consistency of responses to PALS items, we examined the responses to PALS-K tasks from groups defined as Identified and Not-identified for additional instruction under EIRI, based on their PALS-K Summed Score. Since the purpose of PALS-K is to identify children in need of additional instruction, individual items within each task should function differently for Identified and Not-identified groups. For each of the last three mandatory screening windows, this was the case for kindergarten student scores. Table 16 displays the Mantel-Haenszel statistic (based on item scores) for each PALS-K subtask for Spring 2009, Fall 2009, and Spring 2010. As can be seen, the general association statistic is significant for all PALS tasks for both fall and spring.

Table 16 Mantel-Haenszel Statistics (Based on Item Scores) for Identified and Not-identified Groups: Spring 2009, Fall 2009, and Spring 2010

	Spring 2009		Fall 2009		Spring 2010	
PALS Task	GA	<i>p</i>	GA	<i>p</i>	GA	<i>p</i>
Group Rhyme Awareness	19,957	< .001	18,129	< .001	19,890	< .001
Group Beginning Sound Awareness	24,577	< .001	21,567	< .001	27,334	< .001
Alphabet Recognition	20,654	< .001	44,692	< .001	23,275	< .001
Letter Sounds	37,641	< .001	28,192	< .001	39,187	< .001
Spelling	45,584	< .001	16,339	< .001	46,977	< .001
Concept of Word (word list)	32,603	< .001	13,951	< .001	39,704	< .001

*GA= general association

Section VI

Summary

The technical adequacy of PALS-K has been established through pilots and statistical analyses of PALS scores from more than 600,000 kindergarten students statewide over the last ten years. The reliability of individual subtasks is supported through the use of Cronbach's alpha. Reliability coefficients for individual tasks range from .79 to .89 and demonstrate the adequacy of their internal consistency. Inter-rater reliabilities, expressed as Pearson correlation coefficients, have ranged from .96 to .99, indicating that PALS-K tasks can be scored consistently across individuals. In all of these analyses, PALS-K has been shown to be steady, reliable, and consistent among many different groups of users.

Data analyses also support the content, construct, and criterion-related validity of PALS-K. Principal components analyses, discriminant function analyses, receiver-operating characteristic curve analyses, and intercorrelations among tasks provide evidence of the construct validity of PALS-K. Regression analyses have shown the predictive relationship between PALS-K Summed Scores in the fall and Stanford-9 scores in the spring. Coefficients of determination have demonstrated that a significant proportion of the variability in spring Stanford-9 scores can be explained by the PALS-K Summed Score from nine months earlier. Additional evidence of predictive validity is provided by regression equations based on fall PALS-K scores that account for 25% to 45% of the variance in PALS-K and PALS 1–3 scores obtained in the subsequent four screening windows (spring of kindergarten, fall of first grade, spring of first grade, and fall of second grade). Similar analyses have demonstrated the concurrent validity of PALS-K, also using the Stanford-9. In addition, differential item functioning analyses using the Mantel-Haenszel statistic demonstrate the consistency of responses to specific tasks across groups of

Identified and Not-identified students. All of these analyses provide evidence of the validity of PALS-K as an emergent literacy assessment that reliably identifies students in need of additional instruction in reading and writing.

In summary, PALS-K provides an assessment tool with good evidence of validity that can be used reliably to screen students in kindergarten for difficulty in emergent literacy. PALS-K shows evidence of both internal consistency and inter-rater reliability, indicating that it can be administered and scored consistently by different users. PALS-K also shows evidence of content, construct, and criterion-related validity, suggesting that PALS-K indeed captures the underlying constructs associated with emergent literacy.

Section VII

References

- Abouzeid, M. P. (1986). Developmental stages of word knowledge in dyslexia (Doctoral dissertation, University of Virginia, 1986). *Dissertation Abstracts International*, 48, 09A2295.
- Adams, M. (1990). *Beginning to read: Thinking and learning about print*. Cambridge, MA: MIT Press.
- American Educational Research Association (AERA), American Psychological Association (APA), & National Council on Measurement in Education (NCME) (1999). *Standards for educational and psychological testing*. Washington, DC: American Psychological Association.
- Bader, L. (1998). *Bader Reading and Language Inventory* (3rd ed.). Upper Saddle River, NJ: Prentice-Hall.
- Barnes, W. G. (1993). Word sorting: The cultivation of rules for spelling in English. *Reading Psychology: An International Quarterly*, 10, 293-307.
- Bear, D. (1989). Why beginning reading must be word-by-word: Disfluent oral reading and orthographic development. *Visible Language*, 23 (4), 353-367.
- Bear, D., Invernizzi, M., Templeton, S., & Johnston, F. (2004). *Words their way: Word study for phonics, vocabulary and spelling instruction* (3rd ed.). Upper Saddle River, NJ: Merrill.
- Bodrova, E., Leong, D., & Semenov, D. (1999). 100 most frequent words in books for beginning readers. Retrieved from <http://www.mcrel.org/resources/literacy/road/100words>
- Bradley, L., & Bryant, P. (1983). Categorizing sounds and learning to read: A causal connection. *Nature*, 30, 419-421.
- Bradley, L., & Bryant, P. (1985). *Rhyme and reason in reading and spelling*. Ann Arbor, MI: University of Michigan Press.
- Bryant, P., MacLean, M., & Bradley, L. (1990). Rhyme, language, and children's reading. *Applied Psycholinguistics*, 11, 237-252.
- Bryant, P., MacLean, M., Bradley, L., & Crossland, J. (1989). Nursery rhymes, phonological skills and reading. *Journal of Child Language*, 16, 407-428.
- Burton, R., Hill, E., Knowlton, L., & Sutherland, K. (1999). A Reason for Spelling. Retrieved from <http://www.areasonfor.com/index.htm>
- Cantrell, R. (1991). Dialect and spelling in Appalachian first grade children (Doctoral dissertation, University of Virginia, 1991). *Dissertation Abstracts International*, 53, 01A0112.
- Carroll, J. B., Davies, P., & Richman, B. (1971). *The American Heritage word frequency book*. Boston: Houghton Mifflin.
- Catts, H. W. (1993). The relationship between speech-language impairments and reading disabilities. *Journal of Speech & Hearing Research*, 36, 948-958.
- Clay, M. M. (1979). *Reading: The patterning of complex behavior*. Auckland, New Zealand: Heinemann.
- Code of Fair Testing Practices in Education. (1988). Washington, DC: Joint Committee on Testing Practices.
- Dolch, E. W. (1936). *A combined word list*. Boston: Ginn.
- Dorans, N. (1989). Two new approaches to assessing differential item functioning: standardization and the Mantel-Haenszel method. *Applied Measurement in Education*, 2, 217-233.
- EDL (1997). *EDL core vocabularies in reading, mathematics, science, and social studies*. Orlando, FL: Steck-Vaughn.
- Ganske, K. (1999). The developmental spelling analysis: A measure of orthographic knowledge. *Educational Assessment*, 6, 41-70.
- Gill, C. E. (1980). An analysis of spelling errors in French (Doctoral dissertation, University of Virginia, 1980). *Dissertation Abstracts International*, 41, 09A3924.
- Gill, T. P. (1985). The relationship between spelling and word recognition of first, second, and third graders (Doctoral dissertation, University of Virginia, 1985). *Dissertation Abstracts International*, 46, 10A2917.
- Gronlund, N. E. (1985). *Measurement and evaluation in teaching*. New York: Macmillan.
- Harcourt Assessment, Inc. (2004). *Stanford Reading First: Technical data report*. San Antonio, TX: Author.
- 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.
- Huang, F., & Konold, T. (2013). A latent variable investigation of the Phonological Awareness Literacy Screening-Kindergarten assessment: Construct identification and multigroup comparisons between Spanish-speaking English language learners (ELLs) and non-ELL students. *Language Testing*. doi: 10.1177/0265532213496773
- Invernizzi, M. (1992). The vowel and what follows: A phonological frame of orthographic analysis. In S. Templeton & D. Bear (Eds.), *Development of orthographic knowledge and the foundation of literacy* (pp.105-136). Hillsdale, NJ: Lawrence Erlbaum and Associates, Inc.
- Invernizzi, M., Juel, C., Rosemary, C., & Richards, H. (1997). At-risk readers and community volunteers: A three year perspective. *Scientific Studies of Reading*, 1, 277-300.
- Invernizzi, M., Meier, J. D., Swank, L., & Juel, C. (1997). *PALS: Phonological Awareness Literacy Screening*. Charlottesville, VA: University Printing Services.
- Invernizzi, M., Robey, R., & Moon, T. (1999). *Phonological Awareness Literacy Screening (PALS) 1997-1998: Description of sample, first-year results, task analyses, and revisions*. Technical manual and report prepared for the Virginia Department of Education. Charlottesville, VA: University Printing Services.
- Invernizzi, M., Robey, R., & Moon, T. (2000). *Phonological Awareness Literacy Screening (PALS) 1998-1999: Description of sample & second-year results*. Technical manual and report prepared for the Virginia Department of Education. Charlottesville, VA: University Printing Services.
- Johnston, F., Invernizzi, M., & Juel, C. (1998). *Book Buddies: Guidelines for volunteer tutors of emergent and early readers*. New York: Guilford Publications.

- Leslie, L., & Caldwell, J. (1995). *Qualitative Reading Inventory-II*. New York: HarperCollins.
- Lyon, R. (1998). Overview of reading and literacy initiatives. Retrieved January 15, 2002, from National Institute of Health, National Institutes of Child Health and Human Development, Child Development and Behavior Branch Web site: <http://156.40.88.3/publications/pubs/jeffords.htm>
- McBride-Chang, C. (1998). The development of invented spelling. *Early Education & Development*, 9, 147-160.
- McLoughlin, J., & Lewis, R. (2001). *Assessing students with special needs*. Upper Saddle River, NJ: Merrill/Prentice Hall.
- Mehrens, W., & Lehmann, I. (1987). *Using standardized tests in education*. New York: Longman.
- Meltzoff, J. (1998). *Critical thinking about research*. Washington, DC: American Psychological Association.
- Morris, D. (1981). Concept of word: A developmental phenomenon in the beginning reading and writing process. *Language Arts*, 58, 659-668.
- Morris, D. (1992). What constitutes at risk: Screening children for first grade reading intervention. In W. A. Second (Ed.), *Best practices in school-speech pathology* (pp.17-38). San Antonio, TX: Psychological Corporation.
- 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. (1999). *The Howard Street tutoring manual*. New York: Guilford Press.
- Perney, J., Morris, D., & Carter, S. (1997). Factorial and predictive validity of first graders' scores on the Early Reading Screening Instrument. *Psychological Reports*, 81, 207-210.
- Rathvon, N. (2004). *Early reading assessment: A practitioner's handbook*. New York: The Guilford Press.
- Richardson, E., & DiBenedetto, B. (1985). *Decoding skills test*. Los Angeles: Western Psychological Services.
- Rinsland, H. D. (1945). *A basic vocabulary of elementary school children*. New York: Macmillan.
- 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.
- Santa, C., & Hoen, T. (1999). An assessment of early steps: A program for early intervention of reading problems. *Reading Research Quarterly*, 34, 54-79.
- Scarborough, H. S. (1998). Early identification of children at risk for reading disabilities: Phonological awareness and some other promising predictors. In B. K. Shapiro, P. J. Accardo, & A. J. Capute (Eds.), *Specific reading disability: A view of the spectrum* (pp. 75-199). Timonium, MD: York Press.
- Scarborough, H. S. (2000). *Predictive and causal links between language and literacy development: Current knowledge and future direction*. Paper presented at the Workshop on Emergent and Early Literacy: Current Status and Research Direction, Rockville, MD.
- Schlagal, R. C. (1989). Informal and qualitative assessment of spelling. *The Pointer*, 30(2), 37-41.
- Shanker, J. L., & Ekwall, E. E. (2000). *Ekwall/Shanker reading inventory* (4th ed.). Boston: Allyn & Bacon.
- Snow, C., Burns, M., & Griffin, P. (Eds.). (1998). *Preventing reading difficulties in young children*. Washington, DC: National Academy Press.
- Stanford Achievement Test* (9th ed.). (1996). San Antonio, TX: Harcourt Brace.
- Stieglitz, E. (1997). *The Stieglitz informal reading inventory* (2nd ed.). Needham Heights, MA: Allyn & Bacon.
- Superintendents' Memo No. 081-09 (2009). *Early Intervention Reading Initiative — Application Process for the 2009-2010 school year*. Retrieved May 23, 2005, from Commonwealth of Virginia Department of Education Web site: http://www.doe.virginia.gov/info_centers/administrators/superintendents_memos/2009/081-09.shtml
- Swank, L.K. (1991). A two level hypothesis of phonological awareness (Doctoral dissertation, University of Kansas, 1991). *Dissertation Abstracts International*, 53-08A, 2754.
- Swank, L.K. (1997). Linguistic influences on the emergence of written word decoding in first grade. *American Journal of Speech-Language Pathology: A Journal of Clinical Practice*, November 1997.
- Swank, L.K., & Catts, H.W. (1994). Phonological awareness and written word decoding. *Language, Speech, and Hearing Services in Schools*, 25, 9-14.
- Swets, J.A., Dawes, R. M., Monahan, J. (2000). Psychological science can improve diagnostic decisions. *Psychological Science in the Public Interest*, 1, 1-26.
- Templeton, S. (1983). Using the spelling/meaning connection to develop word knowledge in older students. *Journal of Reading*, 27, 8-14.
- Torgesen, J. K., & Davis, C. (1996). Individual differences variables that predict responses to training in phonological awareness. *Journal of Experimental Child Psychology*, 63, 1-21.
- Torgesen, J. K., & Wagner, R. K. (1998). Alternative diagnostic approaches for specific developmental reading disabilities. *Learning Disabilities Research & Practice*, 31, 220-232.
- Torgesen, J. K., Wagner, R. K., Rashotte, C. A., Burgess, S., & Hecht, S. (1997). Contributions of phonological awareness and rapid automatic naming ability to the growth of word-reading skills in second to fifth grade children. *Scientific Studies of Reading*. Vol 1(2), 161-185.
- Vellutino, F. R., Scanlon, D. M., Sipay, E. R., Small, S. G., Pratt, A., Chen, R., & Denckla, M. B. (1996). Cognitive profiles of difficult-to-remediate and readily remediated poor readers: Early intervention as a vehicle for distinguishing between cognitive and experiential deficits as a basic cause of specific reading disability. *Journal of Educational Psychology*, 88, 601-638.
- Viise, N. (1992). A comparison of child and adult spelling (Doctoral dissertation, University of Virginia, 1992). *Dissertation Abstracts International*, 54, 5A1745.
- Virginia Department of Education. (1995). *Standards of learning for Virginia public schools*. Richmond, VA: Commonwealth of Virginia Board of Education.
- Worthy, M. J., & Invernizzi, M. (1990). Spelling errors of normal and disabled students on achievement levels one through four: Instructional implications. *Bulletin of the Orton Society*, 40, 138-151.
- Yopp, H. K. (1988). The validity and reliability of phonemic awareness tests. *Reading Research Quarterly*, 23(2), 159-177.
- Zou, K. H., O'Malley, A. J., & Mauri, L. (2007). Receiver-operating characteristic analysis for evaluating diagnostic tests and predictive models. *Circulation*, 115, 654-657.
- Zutell, J. B. (1975). Spelling strategies of primary school children and their relationship to the Piagetian concept of decentration (Doctoral dissertation, University of Virginia, 1975). *Dissertation Abstracts International*, 36, 8A5030.

Section VIII

Endnotes

¹ Supts. Memo No. 081–09, 2009.

² Virginia Department of Education, 1995, p. 61.

³ Catts, 1993; Lyon, 1998; Scarborough, 1998, 2000; Torgesen & Wagner, 1998; Torgesen, Wagner, Rashotte, Burgess, & Hecht, 1997; Vellutino, Scanlon, Sipay, Small, Pratt, Chen, & Denckla, 1996.

⁴ Swank, 1991; Yopp, 1988.

⁵ Bradley & Bryant, 1983, 1985; Bryant, MacLean, & Bradley, 1990; Bryant, MacLean, Bradley, & Crossland, 1989; Swank, 1997; Swank & Catts, 1994.

⁶ Bradley & Bryant, 1983, 1985; Bryant et al., 1989; Bryant et al., 1990; Swank, 1997; Swank & Catts, 1994.

⁷ Bear, Invernizzi, Templeton, & Johnston, 2004.

⁸ Mehrens & Lehmann, 1987.

⁹ Morris, 1992.

¹⁰ Johnston, Invernizzi, & Juel, 1998.

¹¹ Invernizzi, Juel, Rosemary, & Richards, 1997; Morris, 1999; Perney et al., 1997; Santa & Hoiem, 1999.

¹² Adams, 1990; Snow, Burns, & Griffin, 1998.

¹³ Invernizzi, Meier, Swank, & Juel, 1997.

¹⁴ McBride-Chang, 1998.

¹⁵ Torgesen & Davis, 1996.

¹⁶ Abouzeid, 1986; Barnes, 1993; Bear, 1989; Cantrell, 1991; Ganske, 1999; C. E. Gill, 1980; T. P. Gill, 1985; Henderson, 1990; Henderson & Beers, 1980; Invernizzi, 1992; Schlagal, 1989; Templeton, 1983; Viise, 1992; Worthy & Invernizzi, 1990; Zutell, 1975.

¹⁷ Clay, 1979; Henderson & Beers, 1980; Morris, 1981; Roberts, 1992.

¹⁸ Morris, 1993.

¹⁹ Snow et al., 1998.

²⁰ Bodrova, Leong, & Semenov, 1999.

²¹ Henderson, 1990.

²² Burton, Hill, Knowlton, & Sutherland, 1999.

²³ Dolch, 1936.

²⁴ Rinsland, 1945.

²⁵ Leslie & Caldwell, 1995.

²⁶ Richardson & DiBenedetto, 1985.

²⁷ Shanker & Ekwall, 2000.

²⁸ Johnston et al., 1998.

²⁹ Morris, 1999.

³⁰ Carroll, Davies, & Richman, 1971.

³¹ McLoughlin & Lewis, 2001.

³² Invernizzi, Robey, & Moon, 1999.

³³ Rathvon, 2004, p. 250–261.

³⁴ Invernizzi et al., 1997.

³⁵ Invernizzi, Robey, & Moon, 2000.

³⁶ Mehrens & Lehmann, 1987.

³⁷ Idem.

³⁸ Swank, 1991; Yopp, 1988.

³⁹ Standards for Educational and Psychological Testing, 1999.

⁴⁰ Stanford-9, 1996.

⁴¹ Meltzoff, 1998.

⁴² Gronlund, 1985.

⁴³ Invernizzi et al., 1999.

⁴⁴ Huang & Konold, 2013.

⁴⁵ Swets, Dawes, & Monahan, 2000; Zou, O'Malley, & Marui, 2007

⁴⁶ Harcourt Assessment, Inc., 2004

⁴⁷ See http://www.rti4success.org/tools_charts/popups_screening/scoring/classification.html

⁴⁸ Dorans, 1989.