

TECHNICAL REPORT

**TEXAS PRIMARY READING INVENTORY
(1998 Edition)**

Prepared for the Texas Education Agency

by the

Center for Academic and Reading Skills

**University of Texas-Houston Health Science Center
And
University of Houston**

Barbara R. Foorman

Jack M. Fletcher

David J. Francis

Coleen D. Carlson

Dung-Tsa Chen

Angeliki Mouzaki

Christopher Schatschneider

Kimberly Wristers

Rita H. Taylor

Table of Contents

List of Tables	4
Executive Summary	7
Introduction	11
Screens	11
Inventories	13
Development of the Screens	14
Longitudinal Study	14
Measures	14
Procedures	20
Kindergarten Screens	21
Beginning Grade 1 Screen	27
End Grade 1 Screen	30
Beginning Grade 2 Screen	32
Revision of the Inventory	34
TPRI Field Study	37
Design	37
Procedures	37
Reliability Analyses	38
End of kindergarten analysis	40
Beginning grade 1 analysis	48

End of grade 1 analysis	56
Beginning grade 2 analysis	62
Validity Analyses	68
Content validity	68
Item development	70
Criterion-related concurrent validity	71
End of kindergarten analysis	72
Beginning grade 1 analysis	74
End of grade 1 analysis	75
Beginning grade 2 analysis	77
Field Study Questionnaire: Teacher Opinions	81
Recommended Revisions for Future Development of the TPRI	86
References	88

LIST OF TABLES

Table 1. Median Reliability Coefficients for the TPRI Screens and Inventories	9
Table 2. Time of Assessments, Endpoint Times, and Sample Sizes for Development of Screens	20
Table 3. End of Grade 1 Basic Reading Using December Letter-sounds and Phonological Awareness	22
Table 4. End of Grade 1 Basic Reading Using April Letter-sounds and Phonological Awareness	23
Table 5. Phonological Awareness Item Predictors	25
Table 6. End of Grade 1 Basic Reading Using December Kindergarten Decision Rules	26
Table 7. End of Grade 1 Basic Reading Using April Kindergarten Decision Rules	26
Table 8. Ethnicity Analyses for End of Kindergarten Screening	27
Table 9. End of Grade 1 Reading Using Grade 1 October Word Reading	28
Table 10. End of Grade 1 Reading Using Adjusted Grade 1 October Word Reading	28
Table 11. End of Grade 1 Reading Using Adjusted Grade 1 October Word Reading and Phonological Awareness	29
Table 12. Word reading and phonological awareness items ordered by degree of difficulty.	29

Table 13. Ethnicity Analyses for Beginning Grade 1 Screen	30
Table 14. End of Grade 2 Broad Reading Based on Grade 1 April Word Reading	30
Table 15. End of Grade 2 Broad Reading Based on Adjusted Grade 1 April Word Reading	31
Table 16. End of Grade 2 Broad Reading Based on Adjusted Grade 1 April Word Reading and Phonological Awareness	31
Table 17. Word reading and phonological awareness items for end of Grade 1 screen ordered by difficulty level	32
Table 18. Ethnicity Analyses for End of Grade 1 Screen	32
Table 19. End of Grade 2 Broad Reading Based on October Grade 2 Word Reading	33
Table 20. Word reading items for beginning Grade 2 screen ordered by difficulty level	33
Table 21. Ethnicity Analysis for Beginning Grade 2 Screen	34
Table 22. Design of the TPRI Field Study	38
Table 23. Practical Significance of Reliability Estimates by Magnitude	38
Table 24. Predicted Significance of Item Difficulty Indices by Magnitude	39
Table 25. Clinical Significance for Kappa Estimates of Inter-Rater Reliability	40
Table 26. Reliability Information for the TPRI End of Kindergarten Analysis	43

Table 27. Reliability Information for the TPRI Beginning of First Grade Form	50
Table 28. Reliability Information for the TPRI End of First Grade Form	59
Table 29. Reliability Information for the TPRI Beginning of Second Grade Form	64
Table 30. Concurrent Validity for the TPRI End of Kindergarten Form	73
Table 31. Concurrent Validity for the TPRI Beginning Grade 1 Analysis	76
Table 32. Concurrent Validity for the TPRI End of First Grade Form	78
Table 33. Concurrent Validity for the TPRI Beginning of Second Grade Form	80
Table 34. Teacher responses to the TPRI questionnaire	82

Executive Summary

This report summarizes the technical data underlying the revision and evaluation of the Texas Primary Reading Inventory (TPRI). The TPRI is an English language instrument designed to comply with the requirements of TEC 28.006, which provides school districts the opportunity to evaluate the reading development and comprehension skills of children in kindergarten, grade 1, and grade 2. Recent research indicates that children can be identified as at risk for reading difficulties as early as kindergarten. When reading difficulties are not identified and addressed early in development, they can lead to long-term problems that persist into upper elementary grades and into adulthood (Francis et al., 1996; Juel, 1988). Thus, early identification of reading difficulties is important because effective instruction can eliminate reading difficulties for many children if it is provided early in the child's schooling (Foorman et al., 1998; Torgesen, 1997). The goal of TEC 28.006 is to prevent reading difficulties. The ultimate goal underlying the development and implementation of the TPRI was to enable every child to read and comprehend on or above grade level by the end of Grade 3 and continue to be a successful reader at each subsequent grade. To that end, the TPRI can be used by teachers to identify children at risk for reading difficulties and to further assess the strengths and weaknesses in the reading-related skills of individual children. The TPRI thus facilitates the teacher's ability to set appropriate learning objectives, thereby maximizing each child's opportunities to be a successful reader.

The present revision of the TPRI was completed by the Center for Academic and Reading Skills (CARS) at The University of Texas-Houston Health Science Center and the University of Houston. Through a memorandum of understanding, CARS was asked to revise the TPRI in order to ensure that it was consistent with research on reading skills development and aligned with the Texas Essential Knowledge and Skills (TEKS). CARS was also asked to begin the ongoing process of evaluating the reliability and validity of the TPRI through field studies. CARS was selected for this task because of their long history of research on reading skills development and the technical expertise available through the CARS measurement group at the University of Houston. The report information on the processes used to revise the TPRI, develop screening components, evaluate the reliability and validity of the instrument, and evaluate teacher responses to a pilot implementation. Detailed reports of the reliability and validity of the revised TPRI for the first year of development are provided. The report is divided into four sections: (1) an introduction and description of the revised TPRI; (2) development of the screening assessments; (3) revision of the inventory; and (4) results of a field study to implement the TPRI and evaluate the reliability, validity, and teacher responses to the instrument.

The initial section provides an overview of the screening and inventory portions of the revised TPRI, hereafter referred to as the screen and inventory, respectively. As we explain in the second section, the screens were developed from a longitudinal database collected through a grant from the National Institute of Child Health and Human Development (NICHD; R01 HD28172) to CARS that involved a large sample of children in Kindergarten-Grade 2. It was only through the availability of such a longitudinal database with a large number of children in Kindergarten-Grade 2 that screens with demonstrable predictive validity could be developed. There are five different screens designed for administration in the middle and end of kindergarten, beginning and end of Grade 1, and beginning of Grade 2. All of the screens were developed to minimize the errors that result from not identifying children who need further assistance. Thus, the screens

provide brief assessments (3-5 minutes) that permit identification of children who are not likely at risk for the development of reading difficulties. Children who meet the criteria on the screen do not require administration of the entire inventory, although administration of the comprehension component is encouraged for all children. Teachers should also use their judgment and experience with the child to further evaluate the accuracy of the screening information.

The screen development process was successful. Each of the five screens developed for the TPRI fails to identify less than 10% of the children who end up not reading at expected levels by the end of Grades 1 and 2 (false negative errors). Even when the goal is to identify children who are not at risk, errors involving the over-identification of these children are inevitable (false positive errors). However, errors of this type are viewed as less serious than failing to identify children who are at risk. Moreover, false positive rates were uniformly below 45% for Kindergarten and Grade 1, and dropped dramatically by the second grade (15%).

The screen development section also provides a detailed description of the sample and measures from which the screens were derived, tables summarizing the actual identification rates for each screen, and information on the internal consistency (reliability) and validity of the screening measures. In general, the screens show excellent reliability and validity characteristics and clearly serve the purposes for which they were designed. The use of the screens will save the teacher considerable time and allow more attention to be focused on the children who are struggling with the development of reading skills.

The third section describes the processes used to revise the inventory. There are separate inventories for Kindergarten, Grade 1, and Grade 2. In contrast to the screens, the inventory for a particular grade does not change during the year, and is designed to allow the teacher to evaluate the child up to three times during the year in order to monitor growth and development of reading skills and responses to instructional objectives. The five components of the original TPRI were retained. These components include Book and Print Awareness, Phonemic Awareness, Graphophonemic Knowledge, Reading Accuracy, and Reading Comprehension. Revisions involved rewriting of items, clarifying instructions and procedures for scoring, and redesigning summary sheets. Some new tasks were added and administration procedures for other tasks were modified. The TEA undertook a detailed process of feedback across the state of Texas, with these interactions resulting in additional revisions.

After the revision of the TPRI was completed, CARS conducted a field study of the reliability and validity of the revised TPRI. The goal was to establish tasks that had adequate or better reliability. The fourth section of this manual includes a detailed description of the design of this study. Because there are 5 screens (middle and end of Kindergarten, beginning and end of Grade 1, and beginning of Grade 2), there are five forms of the TPRI with the inventory constant at each grade level. Data was collected on the TPRI forms for the end of kindergarten, beginning of first grade, end of first grade, and beginning of second grade. It was too late in the year (May 1998) to evaluate the form from the middle of kindergarten. Because of the time at which the field study was conducted, children at the end of kindergarten and end of grade 1 were used for the beginning grade 1 and beginning grade 2 assessments, respectively. Students received both the screen and inventory portions of the TPRI on two separate occasions within a one-week

interval. In addition, teachers completed rating scales indicating their perception of the strengths and weaknesses of the TPRI, as well as the training and the opportunity to set learning objectives.

Several indices of reliability and validity were computed. Classical test theory and generalizability theory were used to examine the reliability of the tasks and teacher ratings. Test-retest reliabilities were measured by having the same child evaluated by different teachers over the one-week interval. The use of different teachers produces a lower-bound estimate of test-retest reliability that most likely underestimates the true test-retest reliability characteristics of the instruments. However, using different teachers allows for the evaluation of concurrence between teachers about an individual child, which was a major goal of the field study. In addition, data were provided on the difficulty level of the different tasks and the concurrence between two teachers' ratings of the same child.

The field study consisted of four parts: (1) study design; (2) reliability; (3) validity results; and (4) teacher's opinions. Table 1 summarizes the median reliability coefficients for the four forms of the TPRI that were evaluated. As Table 1 shows, median values for reliability were generally in the good to excellent range for both the screens and inventories.

Table 1. Median Reliability Coefficients for the TPRI Screens and Inventories

Kindergarten – End of the Year Assessment

	Cronbach's Alpha	Generalizability
Screen	.92	.76
Inventory	.89	.70

Grade 1 – Beginning of the Year Assessment

	Cronbach's Alpha	Generalizability
Screen	.86	.86
Inventory	.79	.68

Grade 1 – End of the Year Assessment

	Cronbach's Alpha	Generalizability
Screen	.81	.77
Inventory	.66	.81

Grade 2 – Beginning of the Year Assessment

	Cronbach's Alpha	Generalizability
Screen	.91	.88
Inventory	.67	.76

Other analyses presented in this report showed that difficulty parameters were, with exceptions, at levels in the middle of the ability range for both screens and inventories. In general, items on some tasks were too easy for students in the sample, which reduced reliability estimates. There was more variability in the reliability characteristics of the inventory tasks than in the screens, which was closely linked to difficulty levels. Although specific tasks had weak reliability on certain test forms, the only task for which reliability was consistently less than adequate was Book and Print Awareness in Kindergarten and Grade 1. This seemed to be related to administration difficulties. Kappa estimates of teacher agreement on total task scores were satisfactory. Teacher opinions were uniformly positive, with the most significant concerns involving the need for more training, particularly in using the information provided by the TPRI.

The validity of the TPRI was evaluated in terms of content and construct validity through review of content specifications and through statistical comparisons of the different tasks of the TPRI with independent measures of graphophonemic knowledge, phonological awareness, spelling, rapid serial naming, word reading, and reading comprehension skills. The TPRI has excellent content validity, is clearly consistent with current research on the development of reading skills in children, and is aligned with the TEKS. The pattern of correlations among inventory tasks and the independent measures were statistically significant and in the expected direction, providing support for the validity of the TPRI. Tasks that showed reduced validity were generally those with less than adequate reliability.

The development of the TPRI is an ongoing process and CARS and TEA will continue to refine the TPRI. In the next year, items that demonstrated less than adequate reliability in the field study will be revised and re-piloted. Additional items will be developed that will permit further assessment of reliability characteristics. In addition, further validity studies will be completed. Finally the process of implementation will be monitored and recommendations forwarded to TEA for improving the use of the TPRI in accordance with TEC 28.006.

Introduction

Early reading assessments are required for all children attending public school in Texas in Kindergarten, Grade 1, and Grade 2 through TEC 28.006. The Texas Primary Reading Inventory (TPRI) is an assessment device designed to comply with the requirements of TEC 28.006 by facilitating a teacher's capacity to a) identify children at-risk for reading difficulties in kindergarten, first grade, and second grade; and b) set learning objectives for these at-risk children. Originally developed in 1997 by the English and Language Arts Curriculum Department at the Texas Education Agency (TEA), the Center for Academic and Reading Skills (CARS) at The University of Texas-Houston Health Science Center and the University of Houston was subsequently contracted to revise the TPRI in order to ensure alignment with a) the recently adopted Texas Essential Knowledge and Skills (TEKS); and b) research on reading skills development. In addition, CARS was asked to provide evaluations of the reliability and validity of the TPRI, which will be an ongoing process. The rationale and purposes of the TPRI can be found in the Introduction to each of the three (K, 1, 2) inventories. The present report provides technical information on the development and revision, reliability, and validity of the TPRI at the three grade levels.

At each grade level, the TPRI consists of a screen and an inventory. The *screen* permits the rapid assessment of individual children. Designations of risk status are yielded, which identify children who most likely *do not* need additional assessment. The *inventory* is a detailed assessment of reading and reading-related skills that allows the teacher to gain more in-depth information that can be used to determine the child's level of risk for reading problems. The inventory is primarily designed to help the teacher set learning objectives for the child. Both the screen and the inventory are individually administered and are designed to be given by a trained teacher.

Description of Screens

The screens are based on empirically derived predictors of reading success at the end of Grades 1 and 2. They consist of measures of phonological awareness and word reading skills that predict reading outcomes involving word recognition and comprehension skills in Grades 1-2. Each screen permits the teacher to complete a rapid assessment of all the children in the classroom. They are quick and efficient (less than 10 minutes) and are designed to identify children not at-risk for reading failure who most likely would not need further assessment.

It is important to recognize that a screen is not a diagnostic instrument and that not meeting criteria on the screen does not mean that the child has a reading problem – only a need for closer evaluation. In fact, the screen is designed to identify children who are *not likely* to have reading problems. A child who can meet criteria on the screen is at low risk for reading difficulties. The cut-point on the screen was deliberately set to overidentify children who may be at-risk for reading problems. This decision was made because the failure to identify an at-risk child (false negative error) is more serious than the identification of a child as “at-risk” (false positive error) who experiences no subsequent difficulties with reading.

Whenever a screening device is used, these two kinds of errors will occur. Some children will meet criteria on the screen and yet will subsequently not learn to read at grade level. These errors (false negatives) are the most serious kinds of errors because these children do not receive the additional assistance they require at the earliest possible time, which makes their problems more difficult to remediate at a later time. The TPRI screening component has been developed to minimize the number of false negative errors (under 10%).

Some children will not meet criteria on the screen and yet will be reading successfully by Grade 3, even in the absence of any supplementary assistance. These children represent “false positive” errors and are a concern because they place an unnecessary demand on scarce resources.

The two kinds of errors are tied together; attempts to lower false negative errors lead to an increase in false positive errors, and vice versa. The TPRI screen was designed to keep false positive errors as low as possible, while at the same time minimizing false negative errors. When employing the TPRI, or any other screening device, it is important to keep in mind these two types of errors and to avoid labeling children solely on the basis of their performance on the screen. The screen is not diagnostic; it only serves to indicate the need for further assessment. *Thus, a child who does not meet the screening criteria should be viewed as needing further evaluation. The inventory can be used to reduce or eliminate false positive errors and, if necessary, to plan learning objectives.* Whereas an entire inventory could be given to every child without first administering the screening test, doing so would require substantially more time and effort on the part of teachers. Teachers should also use the child’s progress with the curriculum and their judgment to identify false negatives, i.e., children who meet criteria on the screen, but who struggle with reading. These children should also receive further assessment. It is also recommended that the Comprehension section of the inventories be given to all children, regardless of whether the child meets criteria on the screens, to ensure that no child who may need assistance is not identified.

Even when various phonological awareness and related skills are assessed at the end of kindergarten or beginning of first grade, the link with the development of actual reading skills is not simple. A variety of indirect factors may impinge on the assessment of reading precursor skills that might also produce both false positive and false negative errors. For example, a child may do well on phonological awareness measures because of intense training or extensive literacy experiences, but still struggle with the development of word recognition skills because the training did not include a print component. Similarly, false positive errors may reflect the assessment of children from communities where many families have limited resources and are from diverse cultural and linguistic backgrounds with less exposure to English literacy-related activities. Because of the many possible indirect factors that may influence the child’s development, screening assessments should occur at several points during the child’s development. At a minimum, the following assessment times are recommended: 1) the end of kindergarten; 2) beginning of Grade 1; 3) end of Grade 1, and 4) beginning of Grade 2. The TPRI screen has been developed for each of these timepoints and also in the middle of kindergarten. Screening at the beginning of kindergarten is not encouraged because children need to acclimate to school. There is no screen at the end of Grade 2 because the screen for the

beginning of Grade 2 is highly accurate. Children who do not meet criteria on the beginning Grade 2 screen most likely require intensive intervention and further screening is not indicated.

Description of Inventories

Each inventory is designed to permit a detailed, but informal, assessment of a child's reading and reading-related skills. The primary purpose of the inventories is to assist the teacher in setting learning objectives for the child. By administering the inventory, the teacher may obtain further support for a judgment of the child's reading skills development. Thus, the teacher can set learning objectives for an at-risk child, but also has an opportunity to rectify a false positive error by using the inventory.

Each inventory consists of five components: Book and Print Awareness (Kindergarten and Grade 1 only), Phonemic Awareness, Graphophonemic Knowledge, Reading Accuracy, and Reading Comprehension. The rationale for these components can be found in the Introduction to the TPRI at each grade level.

Inventory items begin at levels that are developmentally less complex (easier) than the screening items. This is because the screens are designed to be more accurate with a child not likely to be "at-risk" for reading problems. In addition, the developmental progression allows the teacher to determine a student's actual level of development in the components of the TPRI.

The inventories are subdivided into different tasks. Each task in the TPRI consists of 5 items. Concepts are described as "developed," when a student provides correct responses to 4 out of 5 questions (3 out of 5 for the Book/Print Awareness task). When a student answers less than 4 questions correctly (3 for the Book/Print Awareness task), then the concept is "still developing", and the teacher can move to the next component of the TPRI. These rules mean that not all items on a task need be administered. Thus, administration time is shortened.

It is recommended that the inventories of the TPRI be administered to students who do not meet criteria on the screen at the beginning, middle, and end of the year so that growth in these literacy-related concepts can be noted and intervention strategies planned. Information gathered from the TPRI can be useful in identifying skill areas in which the student requires further instruction, which is the basis for planning learning objectives. However, the data gleaned from this inventory should be used in conjunction with other evaluations of student performance over time (e.g., samples of student work, lists of books that the student has read, and teacher observations) to provide an in-depth portrait of a student's literacy acquisition. The TPRI was designed to inform instruction rather than to determine placement into programs for students with special needs. The TPRI does not replace the need for further evaluation for special services (e.g., speech and language services). It is also important to be sensitive to students' dialectic, linguistic, and cultural diversity when administering the TPRI. Reliability of scoring can be jeopardized when student and teacher do not share the same dialect. Teachers must be sensitive to the student's dialect. Lack of sensitivity to dialect differences may put the student at risk for reading difficulties because any classroom discussion of word parts (which occurs in beginning reading instruction) will suffer from dialect interference. Flexibility, professional judgment, and knowledge of students should be used in determining errors.

Development of the Screens

Longitudinal Study

It was only possible to develop the screens because of a longitudinal database on the development of reading and reading-related skills made available to the Center for Academic and Reading Skills. This database was accrued in an ethnically and culturally diverse suburban school district in the Houston metropolitan area in 1992-1996 and was supported by a grant from the National Institute of Child Health and Human Development (NICHD; R01 HD28172).

The overall study from which the screens were developed had a modified, longitudinal, time-sequential design in which 945 children in Kindergarten, Grade 1, and Grade 2 were evaluated on reading and reading-related skills four times yearly for one to three years. In addition, individually administered norm-referenced achievement tests were obtained at the end of Grades 1 and 2.

The sample represented a random selection of children in Kindergarten through Grade 2 who participated in a regular education program at three elementary schools. Children were excluded from the sample because of severe emotional problems, uncorrected vision problems, hearing loss, acquired neurological disorders, or classification at the lowest level of English as a second language. The percentage of participation in the Federal Lunch Program in the three schools was 13%, 15%, and 30%. The sample was roughly equally represented by boys and girls. In kindergarten, the following ethnic breakdown was obtained: Caucasian (54%), African-American (18%), Hispanic (15%), Asians (12%), and other (1%). Socioeconomic status was as follows: lower class (9%), working class (43%), and middle class (48%). Slight variations in the sociodemographic characteristics occurred in Grades 1 and 2 for the children selected to participate in this study, but these variations were minor. The sample was culturally and socioeconomically diverse, representing the full range of ethnicities.

Measures

The assessment instruments used to develop the screens included measures used to assess A) intelligence/achievement outcomes in reading, spelling, and math; B) skill growth and development; and C) information on the behavior and environment of the child. Each measure is described in detail below. From these measures, a subset that demonstrated the best predictive validity was selected and used to create the screens.

A. Intelligence/Academic Achievement. These measures included (1) the Wechsler Intelligence Scales for Children-Revised (WISC-R; Wechsler, 1974); (2) tasks 22 and 31 of the Woodcock-Johnson Psychoeducational Test Battery-Revised (WJR; Woodcock & Johnson, 1989) to measure reading decoding; (3) the Spelling test from the Kaufman Test of Educational Achievement (KTEA; Kaufman & Kaufman, 1985); (4) tasks 23 and 32 from the WJR and the comprehension score from the Formal Reading Inventory (Wiederholt, 1986), to measure reading comprehension; and (5) task 24 from the WJR (Calculations) to measure arithmetic proficiency.

1. Wechsler Intelligence Scale for Children-R (WISC-Revised). Each child in the study was administered the WISC-R at the end of Grade 1 and Grade 2. This instrument was standardized on a large sample of children, ages 6.0-16.5 years, stratified for age, gender, race, and SES according to 1970 U.S. census information. Test-retest reliabilities for all tasks ranged from .73-.95. The average correlations among Stanford-Binet IQ scores and WISC-R Verbal, Performance, and Full-Scale IQ's were .71, .60, and .73, respectively (Wechsler, 1974).
2. Decoding skills. Assessment of word recognition ability involves the child's sight word vocabulary. Task 22, "Letter-Word Identification," of the WJR focuses on the ability of children to identify letters and words. Task 31, "Word Attack," examines these same abilities using nonsense words. These are highly reliable measures with internal consistency estimates above .9 and extensive demonstrations of validity (Woodcock & Johnson, 1989).
3. Written spelling. The spelling task of the Kaufman Tests of Educational Achievement was used (Kaufman & Kaufman, 1985) because it has better reliability ($\alpha = .92$) and standardization than other available spelling tests (Moats, 1993). The child spells up to 50 words (in writing) that are orally presented alone and in a sentence. Words are presented in order of increasing difficulty and children who cannot write the words may spell them orally.
4. Reading comprehension. The ability to derive meaning from text was assessed through the reading comprehension tasks of the WJR and the Formal Reading Inventory (FRI; Wiederholt, 1986). Task 23, "Passage Comprehension" ($\alpha = .95$), offers a well-normed measure of silent reading comprehension at the sentence level using a cloze procedure. The subject fills in missing words, relying on what they read for context. Task 32, "Reading Vocabulary" ($\alpha = .97$), assesses understanding of words presented as analogies or opposites.

The Formal Reading Inventory (FRI; Wiederholt, 1986) goes beyond the sentence level and assesses contextual reading comprehension. The FRI consists of four forms, each of which contains 13 passages graded from easiest to hardest. The child reads each passage silently and then answers five comprehension questions. The FRI has internal consistency of .86 and shows good evidence of concurrent validity.

5. Arithmetic. In addition to measures of reading and spelling, the "Calculations" task of the WJR (task 24, $\alpha = .93$) was administered. This is a measure of computational skills in arithmetic, which was important in determining the relative specificity of the predictive validity of the children's growth rates in reading and spelling.

B. Skill Growth and Development. This part of the assessment consisted of measures of phonological processing and memory, semantic language, rapid naming, reading, and spelling. All of these measures have scaling properties that are appropriate for the analysis of change.

1. Phonological processing. In the last two decades a scientific body of evidence has accumulated that identified a phonological processing deficit as a core cause of poor reading (Fletcher et al., 1994; Foorman, Francis, Fletcher, & Lynn, 1996; Stanovich & Siegel, 1994; Wagner, Torgesen, & Rashotte, 1994). In the present study, an early version of the Comprehensive Test of Phonological Processes (CPTT); Wagner, Torgesen, & Rashotte, in press) was used to assess the child's level of phonological awareness. This early version included seven tasks:

a) *Blending onset and rime*. The child is presented with isolated pairs of onsets and rimes at a rate of two per second and asked the child to "put these parts together to make a whole word." There are six practice items and 15 test items, with the number of phonemes in the single-syllable words varying from three to four (e.g., "m-ouse"; "ch-ild").

b) *Blending phonemes into words*. This task is identical to the above task except the child is asked to blend phonemes rather than onsets and rimes. Again, there are 6 practice items and 15 test items (one- and two-syllable words) consisting of two- to six-phonemes (e.g., "i-f"; "w-a-sh"; "j-u-m-p"; "b-a-m-b-oo"; "m-i-s-t-a-ke").

c) *Blending phonemes into nonwords*. This task is identical to the above task except that in place of real words, nonwords are used. Each pseudoword has a parenthetical real word rhyme, which is a near-rhyme provided as a pronunciation key for the experimenter (e.g., "i-th" (with); "y-a-s" (gas); "th-u-ng" (rung); "f-ir-t-u-s" (circus); "n-i-s-p-a-t" (mistake)).

d) *First sound comparison*. In this task, the child is shown a set of pictures and given a target word. The child is asked to point to the picture of the word that begins with the same sound as the target. There are three practice items and 15 test items, consisting of three- to four-phoneme, single-syllable words. For example, in one item the target is "rake" and the alternatives are "ash," "rug," and "see," with the correct response being "rug."

e) *Phoneme elision*. In this task, the child is asked to say a given word and then to say what word would be left if part of the original word was deleted. For example, "Say meat. Now tell me what word would be left if I said meat without saying /t/ ." There are four practice items and 15 test items. All phonemes to be deleted are consonants and all resultant words are real words. The first 12 test items are three-phoneme, single-syllable words where the deletion is at the end of the word for the first 6 items and the beginning of the word for the next 6 items. The last three items are three- to five-phoneme, two-syllable words where the consonant to be deleted is in the middle of the word (i.e., "dri(v)er").

f) *Sound categorization*. This task was adapted from Bradley and Bryant's (1985) sound categorization task. The child is presented with four words and asked to say which one of the words does not sound like the others. For example, in the set "mop-hop-tap-lop," *tap* does not sound like *mop*, *hop*, and *lop*. In each set, the odd word lacks a phoneme shared by the other three words. The missing phoneme is in the middle or end of the word in the first 12 test items and in the beginning of the word in the last three test items. There are four practice items and 15 test items, consisting of three-phoneme, single-syllable words.

g) *Phoneme segmentation*. Children listen to real words and are instructed to "tell me each sound you hear in the word in the order that you hear it." There are four practice items and 15 test items, consisting of two- to five-phoneme, one- and two-syllable words (e.g., *ate, got, jump, person*).

The scores derived from each of these phonological awareness measures represented a summation of the total number of correct responses. Total scores were created in which correct responses were summed. This summation was possible because of the development of an item response model for all the phonological processing measures in this version of the CTPP (Schatschneider et al., 1999).

2. Phonological memory. Children are presented with a non-word and asked to repeat it exactly as presented and as clearly as possible (e.g. 'baf'). There are three practice items followed by up to 25 test items (Wagner et al., in press).

3. Semantic language. The Peabody Picture Vocabulary Test-Revised (PPVT-R; Dunn and Dunn, 1981) was used to measure semantic language skills. The PPVT-R is a well-established measure of recognition vocabulary. In this task, the child is presented with a stimulus word and shown a set of four pictures. The child then chooses the one picture that depicts the word.

4. Rapid Naming. Rapid serial naming of objects, colors, numbers, and letters has been found to be deficient in children with reading disabilities (Denckla & Rudel, 1976; Wolf et al., 1986, 1994). This deficiency has been attributed variously to difficulty in phonological recoding in lexical access (Wagner & Torgesen, 1987) and to deficient orthographic processing ability (Bowers et al, 1994; Wolf et al., 1994). It has also been related to a general rate deficiency (Wolf et al., 1994). McBride-Chang (1996) reported a strong relationship between phonological awareness and rapid naming. Rapid Naming was assessed through administration of Denckla and Rudel's (1976) Rapid Automated Naming (R.A.N.) tests for objects and letters. The objects test consisted of line drawings of common objects (i.e., flag, drum, book, moon, and wagon); the letters task consisted of high-frequency lower-case letters (i.e., a, d, o, s, p). For each task, the stimuli consisted of five items repeated 10 times in random sequences. The child was asked to name each stimulus as quickly as possible. The correct number of responses named within 60 seconds was recorded. Test-retest reliability has been estimated at .57 from kindergarten to Grade 1, which may reflect variability in true change over this age range, and at .77 from Grade 1 to 2 (Wolf et al., 1986).

5. Alphabetic knowledge. Kindergarteners' knowledge of letter names and sounds of the alphabet was assessed by showing the children printed cards with both the upper and lower case letters and asking the child to name the letter and then to say the sound of the letter. Credit was given for no more than one correct sound per letter. Schwa sounds did not receive credit. In kindergarten, measures of this sort are highly predictive of subsequent reading achievement (e.g., Vellutino et al., 1996).

6. Word reading and spelling. To assess growth in word reading, children individually read aloud 50 words aloud that are presented one at a time on 4 x 6 cards. The words are matched for frequency and consistency (Carroll, Davies, & Richman, 1971), are representative of a diversity of linguistic features, and span first through third grade difficulty levels. Scores were based on the number of words read aloud correctly out of 50. Growth in spelling was measured by a dictation test using the word list described above. Analyses were based on item response models that were conducted on these reading and spelling measures.

7. Word reading efficiency test (Torgesen & Wagner, in press). This is a measure of speeded reading of single words. The child is given a list of 104 words divided into 4 columns of 26 words each, and asked to read them as fast as possible. A short (8-word) practice list is presented first. The number of words read correctly and number of errors made, within 45 seconds, is recorded. The child is then given a second list to read, following the same directions. Accuracy (total number of words read correctly on the two lists) divided by time to read the list (45 second limit) was the primary variable.

8. Visual-motor integration. The Beery Test of Visual-Motor Integration (VMI; Beery, 1989) consists of 24 geometric line drawings of increasing complexity and assesses visual-motor integration and graphomotor skills. The drawings must be copied without erasures from a basal of three passes to a ceiling of three failures. Explicit scoring criteria and normative data on 2-15 year olds are provided in the VMI manual. Inter-rater reliability for the Beery is .93; median split-half reliability is .79.

9. Recognition-Discrimination. The Recognition-Discrimination test (Satz & Fletcher, 1982) is a visual-perceptual (matching to sample) task requiring the child to identify a geometric stimulus design among a group of four figures, three of which were rotated and one similar in shape to the stimulus figure. The test has 3 practice items 24 test items, and is timed. This test is one of the tasks in the Florida Kindergarten Battery (Satz & Fletcher, 1982). We included it here as an additional non-linguistic measure because it is motor-free, has good reliability (Kuder-Richardson coefficient of .94), and good predictive validity for reading group classification throughout elementary school (Satz et al., 1978).

C. Behavior and Environmental Information Battery

1. Yale Children's Inventory (YCI; Shaywitz et al., 1992). The YCI is a comprehensive rating scale filled out by parents that evaluates aspects of the child's behavioral adjustment at home. The YCI scales include 74 questions coded 1 (never) to 4 (often) that permit assessment of the child's behavior in several domains. The YCI has been factor analyzed into 11 scales: Attention, Activity, Impulsivity, Tractability, Habituation, Conduct Disorder (CD)-Socialized, (CD)-Aggressive, Negative Affect, and Fine Motor. A second order factor analysis of the 11 scales yielded a two-factor solution consisting of behavioral and cognitive factors. Therefore, the YCI can be represented by 11 narrow band scales and two broad band scales. Extensive reliability and validity analyses are available, with high internal consistency for the scales (.81-.96) and excellent validity (Shaywitz et al., 1992).

2. Multigrade Inventory for Teachers (MIT; Agronin et al., 1992). Evidence from several lines of investigation suggests that the teacher is a good predictor of a child's later school performance. The teacher has developed an internal norm against which to compare students and has the advantage of being able to observe the child over a period of time, considering both good and bad days, and a variety of situations, academic and social. Several investigators have noted that the addition of teacher observations to a screening procedure increases the accuracy significantly. Acknowledging that the observations of the teacher represent an often under-appreciated and under-utilized source of information, the MIT (Agronin et al., 1992) provides for the systematic recording of the teacher's insights in a manner that is consistent for all children. Thus, the MIT provides a mechanism for the child's classroom teacher to record his/her observations on a rating scale that includes precise descriptions of a full range of behavioral styles reflecting the child's processing capabilities, adaptability, behavior, language, fine motor and academic proficiency. The MIT consists of 60 items coded by the teacher on a 0 (never) to 4 (often) scale. Instrument development has resulted in six scales: Academic, Activity, Language, Dexterity, Behavior, and Attention. Each scale is highly reliable, with internal consistencies ranging from .80-.95. The validity of the scales has been demonstrated in several studies (Agronin et al., 1992).
3. Henderson Environmental Learning Process Scales (HELPS; Henderson et al., 1972). The HELPS was developed to measure characteristics of the home, which are related to scholastic achievement. The items involve the family's educational aspirations, interest in and access to printed materials, willingness to involve the children in learning situations outside of school, and parental pursuit of educationally related activities. It is a self-administered questionnaire containing 51 Likert-type items, with reliability of .80. Satisfactory validity is apparent in a range of applications to different clinical and ethnic groups.
4. Family Resource Scale (FRS; Dunst & Leet, 1987). The FRS is an objective measure for assessing the adequacy of family resources and needs. It includes 30 items that measure the perceived adequacy of physical and human resources. Each item is rated on a five point scale ranging from "not at all adequate" to "almost always adequate." As such, the FRS is an important supplement to traditional SES information. Internal consistency is .92; test-retest reliability is .70. Several studies support the validity of the instrument (Dunst & Leet, 1987).
5. Harter Perceived Competence Scale. The Perceived Competence Scale distinguishes five separate domains: scholastic competence, athletic competence, social acceptance, physical appearance and behavior. A "structured alternative format" is used to minimize the likelihood of the child making a socially desirable response. For each item children must first decide whether the statement is true or not true for themselves and then decide whether the statement is "sort of true/not true" or "very true/not true" of themselves. Each item is scored 1 to 4. The scales are

psychometrically sound with reliability (coefficient alpha) exceeding .75 and test-retest reliability for subscales above .80 (Harter, 1982).

6. End of Year Evaluation (EYE). The EYE was developed to enable investigators to systematically collect information reflecting the child's educational placement and experiences. In addition, information is collected on health or attendance factors that might influence school performance. Specifically, the EYE asks about any special services received by the child, and recommendations for the next class placement and recommendations for special services. It also includes information on types of service (i.e., pre-referral, special education, academic or behavioral), age at initiation of services, individual or group placements, and the amount of time spent receiving each service. Grades, absences, tardies, and the results of hearing and visual screening are also recorded.

Procedures

Five screening measures were developed that varied depending on when they were administered. This type of development was possible because children in the parent study were seen four times yearly (October, December, February, April). The sample represented all of the children available at a particular timepoint who also had one of the criterion outcome assessments. Table 2 presents the sample sizes used to relate each of the five screening and outcome assessments, and indicates when the screening and outcome assessments were conducted.

Table 2.
Time of Assessments, Endpoint Times, and Sample Sizes for Development of Screens.

<u>Timepoint</u>	<u>Outcome</u>	<u>Sample Size</u>
1. Middle Kindergarten	End of Grade 1	421
2. End of Kindergarten	End of Grade 1	421
3. Beginning of Grade 1	End of Grade 1	599
4. End of Grade 1	End of Grade 2	376
5. Beginning of Grade 2	End of Grade 2	540

In developing the screening battery, data from the longitudinal project were evaluated to identify the smallest possible subset of measures that could be used to gain the highest levels of predictive validity relative to reading outcomes at the end of Grades 1 and 2. The measures described above were originally selected because of their hypothesized sensitivity to the long-term development of reading skills. The criterion measures were derived from the individually administered achievement tests described above. In order to develop the screening battery, some criteria for end of year performance had to be established. For predictions in kindergarten, we attempted to predict outcomes using the Woodcock-Johnson PsychoEducational Test Battery-Revised (WJR) Basic Reading cluster, which includes measures of word recognition and word attack skills. For first and second graders, the WJR Broad Reading cluster was employed, which includes measures of word recognition skills and a clozed-based reading comprehension

measure. This composite thus incorporates both word recognition and comprehension skills, with the latter particularly essential with the older children.

To define outcomes, or risk for reading failure, a criterion of 0.5 grade equivalents below grade level was used on the outcome measures at the end of Grades 1 and 2. For first graders, risk represented a reading grade equivalent to 1.4 or lower; for second graders, risk represented a grade equivalent of 2.4 or lower. In Grade 1, this corresponded to a score that would be at the 22nd percentile for Basic Reading and 18th percentile for Broad Reading. In second grade, a 2.4 grade equivalent score corresponded to the 35th percentile. The cut-point was deliberately set higher in Grade 2 because of greater stability in the prediction equations and reduction in the amount of time for a child to read at grade level by the third grade. These designations are arbitrary, but correspond to the stated goal of TEC 28.006, which is to have every child reading at grade level by the end of the third grade.

Separate analyses were made of the data available at each of the five timepoints. The general approach was to establish a series of prediction equations that helped identify those variables that contributed uniquely to the prediction of risk status. It is important to recognize that decisions about effective predictors were not based solely on the strength of the relationship between the predictor and criterion measures (e.g., correlation coefficient). Rather, the decision was made on the basis of the accuracy of individual child predictions, as reflected in the rate of both false positive and false negative errors.

Kindergarten Screens

The kindergarten screens were developed using performance in December and April from the longitudinal database (Table 2). December was selected because of the high probability that many children with fewer literacy experiences would do poorly on a kindergarten screen administered at the beginning of the year. In addition, children in kindergarten need time to acclimate to the school environment. Consequently, the earliest time at which a child should be assessed with the TPRI is in the middle of kindergarten. We also selected the end of kindergarten in order to help the teacher identify children who would benefit from administration of the inventory in order to plan learning objectives for the summer and following year.

From the battery of tests described above, five Kindergarten measures were initially selected on the basis of previous research with this data set. This research has established predictive relationships for many of the measures with outcomes at the end of first and second grade. The predictors included measures of letter names, letter sounds, phonological awareness, rapid serial naming (letters), and the Peabody Picture Vocabulary Test-Revised. Other measures in the battery had been previously evaluated and did not show independent contributions to reading outcomes.

The first step in establishing the best set of predictors involved an examination of all possible combinations of the five predictors in predicting outcomes at the end of the Grade. For each possible combination, a linear discriminant function analysis was conducted. We examined both the squared canonical correlation, an index of the strength of the relationship between the predictor and outcome variable(s), and the identification matrices resulting from predicting

outcomes on a case-by-case basis. A variable or set of variables was selected if it exhibited both a) a high squared canonical correlation and b) relatively low numbers of both false positive and false negative errors. In all instances, the prediction set that provided the best identifications using the least number of predictors was selected. Once a set of predictors was selected, a cut-point from the equation expressing the relationship of the predictors and outcomes was established. This cut-point was achieved by deliberately adjusting the equation to establish the lowest possible false positive error rate, while also keeping false negative error rates below 10%.

Regardless of the combination of variables, two kindergarten measures consistently provided the best predictive discrimination: letter-sound naming and phonological awareness. It should be noted that alternative outcome measures were also examined, namely other measures of word recognition skills. The results did not vary according to how outcomes were measured. In addition, predictions to the beginning of Grade 1 were also explored to ensure that excessive instability was not being introduced by the length of time from predictor to criterion. However, results were virtually identical for beginning and end of Grade 1 predictions, so that the beginning Grade 1 analyses (prediction from December of Kindergarten to October of Grade 1) were dropped. Tables 3 and 4 summarize the identification tables for December and April of Kindergarten relative to end of Grade 1 outcomes using the measures of letter-sound naming and phonological awareness.

Table 3.
End of Grade 1 Basic Reading Using December Letter-sounds and Phonological Awareness

Outcomes	Identifications		Total	
	No Risk	At Risk		
No Risk	204	121	325	False positive rate = .37
At Risk	9	87	96	False negative rate = .09
Total	213	208	421	

Table 4.
End of Grade 1 Basic Reading Using April Letter-sounds and Phonological Awareness

Outcomes	Identifications		Total	
	No Risk	At Risk		
No Risk	218	106	324	False positive rate = .33
At Risk	11	86	97	False negative rate = .11
Total	229	192	421	

The next step involved determining the specific items for both the letter-sound identification and phonological awareness measures that would produce adequate identification (or identification similar to that of the linear discriminant analyses using the full item set from a measure), and determining the cut-off points for the scores on the sets of items selected. Cut-offs that produced the most desirable classification were selected. This next step permitted further reduction in the amount of time required to complete the screen and was only possible because of the extensive item-level research completed on these instruments (Schatsneider et al., 1999).

Letter-sound identification

Knowledge of letter sounds is actually a rudimentary form of phonological awareness. There are only 26 letters in the English alphabet, so that letter-sounds represents a finite universe of items that cannot be expanded by creation of new items. The letter sounds also display a clear ordering with respect to difficulty (Treiman et al., in press). Not all sounds are equally predictive, so in order to score above the cut-point determined by the discriminant analysis, children would have to be successful on at least some of the most difficult letter-sound items. Thus, it is possible that an optimal subset of items for the screen could be constructed from the 13 most difficult letter-sound items. We therefore evaluated a screen that consisted of less than half of the original items, but which yielded comparable discrimination to the full item set. The letter-sound items to be used were determined by examining the relative frequency of incorrect responses on each letter. These percent incorrect rates were examined across December and April for the entire sample, and items that had the highest error rates across waves were selected.

This procedure allowed for the determination of the 10 most difficult items across both waves. The items were similar, but the order varied somewhat over time. These are the most difficult 10 items (listed in order of easiest to hardest):

December: N L O E I Q W X U Y April: L O N W E I Q U Y X

The letters Q and X present some difficulty from a linguistic perspective because neither letter has a clearly identifiable sound in isolation in English. The next most difficult items are R and H, which are not appreciably easier than L, O, and N. Thus, the decision was made to substitute R and H for X and Q, resulting in the following list of 10 items:

L O N I R E H W U Y

This list has a reliability (coefficient alpha) of .90 and a bivariate correlation with end of Grade 1 Basic Reading of .51 (December) and .54 (April).

The cut-point for the 10 items was adjusted to account for the developmental progression in the two timepoints, which resulted in the following decision rules:

Decision Rule Middle KG: Less than 4 out of 10 correct
Decision Rule End KG: Less than 8 out of 10 correct

Phonological Awareness

To evaluate the contribution of the individual phonological awareness items to predictions, the total number correct out of the 10 letter-sound items was plotted against the phonological awareness score to determine the most appropriate cut-point on phonological awareness. By manipulating the cut-points on phonological awareness, it was determined that the best discrimination resulted from the following decision rules:

December: Letter-sounds = less than 4 of 10 and phonological theta of $-.80$

April: Letter-sounds = less than 8 of 10 and phonological theta of $-.37$

The next step involved the assembly of brief lists of phonemic awareness items that maximally discriminated around the item level cut-points given above and to determine the number correct scores on the item sets that optimally identified students as above or below the item level cut-point. Because the phonological theta scores are on the same scale as item difficulties, the selection of items is a straightforward process.

The eight items with difficulty parameters closest to the cut-off point were selected for each wave. These items are summarized in Table 5, and result in scales that have reliabilities (coefficient alpha) of .91 for both December and April. The bivariate correlations with end of Grade 1 Basic Reading are .50 (December) and .48 (April).

Table 5.
Phonological Awareness Item Predictors

	<u>December</u>	
P-ICK		M-ARK
F-IGHT		CH-IN
TH-ANK		S-AW
R-A-SH		W-I-SH
	<u>April</u>	
W-ILL		S-OO-N
L-A-S-T		I-SH
V-AW		F-OO
W-OY		H-A-SS

The total number of correct out of the 10 letter-sound items was plotted against the total number of phonological items correct out of the 8 selected for each wave. Cut-off points on the phonological sum variable were manipulated and it was determined that the best discrimination resulted from the following decision rules:

Letter-sounds:

Decision Rule Middle KG: Less than 4 out of 10 correct

Decision Rule End KG: Less than 8 out of 10 correct

If the child did not meet these criteria, the phonological awareness items are administered and the following rule applies:

Phonological Awareness:

Decision Rule Middle and End KG: Less than 6 out of 8 correct

The classification tables for the final cut-points are presented in Tables 6 and 7.

Table 6.
End of Grade 1 Basic Reading Using December Kindergarten Decision Rules

Outcomes	Identifications		Total	
	No Risk	At Risk		
No Risk	181	143	324	False positive rate = .44
At Risk	5	92	97	False negative rate = .05
Total	186	235	421	

Table 7.
End of Grade 1 Basic Reading Using April Kindergarten Decision Rules

Outcomes	Identifications		Total	
	No Risk	At Risk		
No Risk	200	124	324	False positive rate = .38
At Risk	10	87	97	False negative rate = .10
Total	210	211	421	

Additional Analyses

To ensure equivalence of discrimination, the criterion functions were examined across ethnic groups. In addition, the Multi-Grade Inventory for Teachers (MIT), a teacher rating scale, and student ages were examined to determine whether identification improved with the addition of any of these student characteristics.

To examine whether there was differential accuracy of the prediction equation across ethnic groups, a variable representing the identifications produced from the cut-points for letter-sounds and phonological awareness was constructed. This variable was then used in conjunction with ethnicity to determine whether outcomes (risk, no risk) varied according to these two factors or their interaction. In both December and April, there was a significant interaction for ethnicity by predicted classification. However, post-hoc examination indicated that classification of risk status was more accurate for Hispanic and African-American than Caucasian students, so that bias due to ethnic differences was not in the direction of concern. Table 8 presents the results of this analysis for the end of Kindergarten screen; the results for the middle Kindergarten screen were similar.

Table 8.
Ethnicity Analyses for End of Kindergarten Screening

Factor	DF	Type III SS	Mean Square	F Value	Pr > F
Identifications	1	6.77	6.77	47.24	0.0001
Ethnicity	3	1.38	0.46	3.21	0.02
Identifications*Ethnicity	3	1.07	0.35	2.49	0.06

Following a similar process, while some MIT scales slightly improved identification, none provided enough improvement above and beyond the two primary predictors to justify the added degree of difficulty of administration. In addition, age did not significantly improve identification relative to the primary screening measures of letter-sounds and phonological awareness.

Beginning Grade 1 Screen

The Grade 1 and Grade 2 screens followed the same development process as the kindergarten screens. Consequently, the results for Grade 1 and Grade 2 screens will be summarized with less detail. Using the same procedures as in Kindergarten to select the best combination of predictors, the Beginning Grade 1 screen consisted of three measures used to predict End of Grade 1 Broad Reading:

- 1) Letter-sound task - the child is asked to identify the sounds of 10 letters (with a cut-off of 8 out of 10)
- 2) Word reading task - the child is asked to read 10 words (with a cut-off of 8 out of 10)
- 3) Phonological awareness task - the child is asked to do six phoneme blending tasks (with a cut off of 5 out of 6)

The decision to include the letter-sound task was based on conceptual reasons. First, since kindergarten is not mandatory in the state of Texas, it was felt that some continuity with the End of Kindergarten screen should be maintained. The letter-sound task in the Beginning of Grade 1 screen is identical to the letter-sound task in the End of Kindergarten screen. Since the identification of letter-sounds is the easiest of the three tasks in the Beginning of Grade 1 screen, children who could not meet criteria on the letter-sound screen at the beginning of Grade 1 should receive the TPRI inventory for Grade 1.

The items selected for the word reading task were based on a discriminant function analysis of 599 children who had both October and May data for Grade 1 (Table 9). After examining a number of different models (as in the kindergarten screen), we determined that the best predictor was the 50 item measure of Word Reading described above. The score used in the analysis is derived from an Item Response Theory (IRT) model for the 50 items. The score is an estimate of a student's word reading ability and is referred to as "theta" score (w_{θ}). This measure is scaled to have a mean of zero and a standard deviation of 1 in Grade 2.

Table 9.
End of Grade 1 Reading Using Grade 1 October Word Reading

Outcomes	Identifications			
	No Risk	At Risk	Total	
No Risk	309	171	480	False positive rate = .36
At Risk	13	106	119	False negative rate = .11
Total	322	277	599	

From this analysis, the identification rule was manipulated so that the percentage of false negatives would fall below 10% (Table 10). This criterion of accuracy was met when a cut-off of -1.07 on w_theta was used.

Table 10.
End of Grade 1 Reading Using Adjusted Grade 1 October Word Reading

Outcomes	Identifications			
	No Risk	At Risk	Total	
No Risk	279	201	480	False positive rate = .42
At Risk	6	113	119	False negative rate = .05
Total	285	314	599	

While this cut-off achieved the desired accuracy rate for false negative errors, the number of false positives was higher than we felt was desirable. To reduce the false positive error rate, we employed a classification and regression tree (CART) analysis. Using CART, we determined that the false positive rate could be further reduced by using phonological awareness scores. Using a decision rule that identified children as being at risk if they scored below -1.07 on w_theta and below 0.0 on p_theta , the following identification table was obtained:

Table 11.
End of Grade 1 Reading Using Adjusted Grade 1 October Word Reading and Phonological Awareness

Outcomes	Identifications			
	No Risk	At Risk	Total	
No Risk	305	175	480	False positive rate = .37
At Risk	8	111	119	False negative rate = .07
Total	313	286	599	

Eight items were selected from the Word Reading list and 6 items were selected from the phonological awareness measure with difficulty parameters that were nearest to the respective cutpoints for the two measures. Using the selected items, number correct cut-offs were established for each measure and the identification accuracy of the resulting screening instrument and decision rule was re-examined. The word reading and phonological awareness items selected are presented in Table 12:

Table 12.
Word reading and phonological awareness items ordered by expected degree of difficulty.

<u>Word Reading</u>	<u>Phonological Awareness</u>
<u>Item</u>	<u>Item</u>
1) GOLD	1) H-A-S
2) BEEN	2) S-OO-N
3) SUMMER	3) F-A-S-P
4) CAMP	4) S-UR-F-A-C-E
5) FRIEND	5) SH-U-NG
6) BECOME	6) TH-U-N-T
7) WASH	
8) RICH	

Reliability (coefficient alpha) was estimated at .90 for the word reading list and .91 for the phonological awareness items. Bivariate correlations with the end of Grade 1 WJR Broad Reading cluster score were .81 and .67, respectively.

The screening battery was also analyzed to see if the prediction equations performed differently among the four ethnic groups (African-American, White, Asian, and Hispanic). Using the same procedure as with the kindergarten ethnicity analyses, we found that the prediction equations did not perform significantly different for any of the groups (Table 13). This would be indicated by a significant ethnicity effect or interaction involving ethnicity.

Table 13.
Ethnicity Analyses for Beginning Grade 1 Screen

Factor	DF	Type III SS	Mean Square	F Value	Pr > F
Identification	1	10.64	10.64	82.24	0.0001
Ethnicity	3	0.43	0.14	1.11	0.35
Identification*Ethnicity	3	0.26	0.09	0.66	0.58

End of Grade 1 Screen

The TPRI End of Grade 1 screen consisted of two measures predicting End of Grade 2 Broad Reading:

- 1) Word reading task - the child is asked to read 10 words (with a cut-off of 8 out of 10)
- 2) Phonological awareness task - the child is asked to do 6 phoneme blending tasks (with a cut off of 5 out of 6)

The items on the word reading list were selected based on a discriminant function analysis (Table 14) of 376 children who had both data from Grade 1 Spring and end of Grade 2 (WJR Broad Reading score). After analyzing a number of different models, it was determined that the best model included just the Word Reading theta score (w_{θ}).

Table 14.
End of Grade 2 Broad Reading Based on Grade 1 April Word Reading

Outcomes	Identifications		Total	
	No Risk	At Risk		
No Risk	254	45	299	False positive rate = .16
At Risk	12	65	77	False negative rate = .15
Total	266	110	376	

From this analysis, we adjusted the classification rule so that the percentage of false negatives would fall below 10%. When the cut-off is set at w_{θ} equal to $-.2$, the number of false negatives drops below 10% (Table 15).

Table 15.
End of Grade 2 Broad Reading Based on Adjusted Grade 1 April Word Reading

Outcomes	Identifications		Total	
	No Risk	At Risk		
No Risk	217	82	299	False positive rate = .27
At Risk	3	74	77	False negative rate = .04
Total	220	156	376	

With the number of false negatives reduced, we then tried to reduce the false positive rate using CART. We determined that we could reduce our false positives by using the children's phonological awareness that score (p_{θ}). Using a decision rule that identified children as being at risk if they score below -0.2 on w_{θ} and below $.80$ on p_{θ} , we obtained the identification table in Table 16:

Table 16.
End of Grade 2 Broad Reading Based on Adjusted Grade 1 April Word Reading and Phonological Awareness

Outcomes	Identifications		Total	
	No Risk	At Risk		
No Risk	229	70	299	False positive rate = .23
At Risk	6	71	77	False negative rate = .08
Total	235	141	376	

Eight word reading items and 6 phonological awareness items with difficulty parameters that were nearest to their respective cutpoints were then selected.

Table 17.
Word reading and phonological awareness items for end of Grade 1 screen ordered by expected difficulty level.

<u>Word Reading</u>	<u>Phonological Awareness</u>
<u>Item</u>	<u>Item</u>
1) UNHAPPY	1) V-IR-T-U-S
2) BLOCK	2) Z-E-LL-O
3) HAIR	3) SH-L-UE
4) MATTER	4) A-R-O-U-N-D
5) SOMEWHERE	5) SH-A-M-P-OO
6) TABLE	6) L-E-S-T-A-T
7) YOUNG	
8) SHOES	

The word list has reliability (coefficient alpha) of .92 and a bivariate correlation with end of Grade 2 Woodcock-Johnson Broad Reading of .82. The phonological awareness scale had an estimated reliability (coefficient alpha) of .92 and a bivariate correlation with end of Grade 2 Broad Reading of .60.

The performance of this prediction equation across the 4 ethnic groups was also analyzed to see if the prediction equations performed equally well for all groups (African-American, White, Asian, and Hispanic). There were no significant effects or interactions involving ethnicity.

Table 18.
Ethnicity Analyses for End of Grade 1 Screen

Factor	DF	Type III SS	Mean Square	F Value	Pr>F
Identifications	1	11.57	11.57	104.79	0.0001
Ethnicity	3	0.23	0.08	0.70	0.55
Identifications*Ethnicity	3	0.10	0.03	0.03	0.83

Beginning Grade 2 Screen

The Beginning Grade 2 screen consisted of a word reading task in which the child read 10 words (with a cut-off of 8 out of 10). No other variables uniquely contributed to outcomes at the beginning of Grade 2. As before, the items selected for the word reading task were based on a discriminant function analysis of 537 children who had both beginning and end of second grade outcome data. After analyzing a number of different models, it was determined that the best model included just the Word Reading theta score (w_theta). The results are presented in Table 19.

Table 19.
End of Grade 2 Broad Reading Based on October Grade 2 Word Reading

Outcomes	Identifications		Total	
	No Risk	At Risk		
At Risk	370	65	435	False positive rate = .15
No Risk	9	93	102	False negative rate = .09
Total	379	158	537	

From this analysis, we determined that no adjustment of the cut-point was necessary. From the canonical variate, we calculated a cut-point of $-.03$ on w_{θ} . Using CART, we determined that we could not reduce the false positive rate substantially while also keeping our percentage of false negatives below 10%. This finding provided further support for a model involving only w_{θ} .

Table 20 lists the word reading items and their respective difficulty parameters.

Table 20.
Word reading items for beginning Grade 2 screen ordered by difficulty level.

<u>Word Reading</u>	
<u>Item</u>	<u>Difficulty Level</u>
1) QUEEN	-1.09
2) COLORFUL	-1.00
3) MONKEYS	-0.96
4) SHOES	-0.93
5) SWING	-0.90
6) SHELF	-0.87
7) BEGUN	-0.75
8) OCEAN	-0.68

This set of items has a reliability (coefficient alpha) of $.86$ and a bivariate correlation with end of Grade 2 Woodcock-Johnson Broad Reading of $.80$.

Again, the possibility of differential prediction for ethnic groups was examined (Table 21). No evidence was found to indicate that this identification rule differentially predicted the criterion across the ethnic groups.

Table 21.

Ethnicity Analysis for Beginning Grade 2 Screen

Factor	DF	Type III SS	Mean Square	F Value	Pr > F
Identifications	1	24.05	24.05	276.33	0.0001
Ethnicity	3	0.10	0.03	0.37	0.77
Identification*Ethnicity	3	0.54	0.18	2.08	0.10

Revision of the Inventory

Revision Process

The inventory was revised to enhance alignment with both the Texas Essential Knowledge and Skills (TEKS) and with current research on the development of reading skills in children. The five components of the Texas Primary Reading Inventory were retained. Revisions were made to clarify instructions and scoring, along with revisions to the Summary Sheet. Initial revisions were made by the staff at the Center for Academic and Reading Skills based on responses gathered from teachers who used the original TPRI. The revised inventory was then presented at a meeting of the Regional Service Centers in January, 1998, and feedback was solicited. In addition, individuals associated with literacy-related issues in the state of Texas were consulted about the TPRI, and a formal meeting was held in March, 1998. Finally, a large group of teachers and assessment professionals was trained to administer the TPRI and these individuals provided feedback. The following is a summary of the major revisions to the TPRI organized by the five domains subsumed within the inventory.

The format of the revised inventory was initially very similar to that of the original TPRI. After a review of all feedback, it was necessary to include extensive administration guidelines for the tasks that required more advanced assessment skills. Furthermore, because of the twofold nature of the instrument (screening and inventory) and the number of tasks included, it was determined that an extensive formatting revision was necessary. Accordingly, specific directions for scoring and further assessment were clearly provided on each page and a summary sheet was modified to include the information/score obtained.

Book and print awareness

Objective: It is important to determine what the student knows about books, print and language, and the reading process.

Revision: Questions in this section were based on the corresponding section from the old version. However, questions regarding the author and illustrator of the book were excluded because of their limited value in estimating the reading ability of children with limited literacy experiences. The questions were further modified for clarity based on feedback received at both the January and March meetings.

Phonological awareness

Objective: Research has shown that it is essential for a child to have the ability to detect internal sound structures within spoken language in order to begin to learn sound-symbol relationships.

Revisions: Items were selected by CARS staff based on item response analysis of the assessment instruments used in CARS research in order to determine level of difficulty among different phonological awareness tasks (Schatschneider et al., 1999). Tasks in the phonological awareness section of the inventory were arranged in increasing order of difficulty. An effort was made to use real words instead of the pseudowords used in the previous version. Some modifications were made to address concerns from the January and March meetings with respect to linguistic properties of the letters/sounds included in the test items (i.e. avoid use of stop consonants in some items).

Graphophonemic knowledge

Objective: Graphophonemic knowledge is the recognition of letters and the understanding of sound-symbol relationships. When phonological awareness begins to develop, the student's understanding of letters and sound-symbol correspondences should be determined. In addition, the student is asked on the TPRI to link their phonological awareness to letter-sound knowledge, and to apply their knowledge of sound-symbol relationships to encode regularly spelled words, increasingly complex spelling patterns, orthographic conventions, and morphological units such as plurals and verb tenses.

Revision: Items were developed to provide a model of good instructional practices for the teacher. A word building approach was adapted for the Graphophonemic sections for Grade 1 and 2. With this approach, students are asked to substitute initial and final consonants, middle vowels, and initial and final blends by directly manipulating magnetic or cardboard letters. After several administrations of this section of the inventory, it became apparent that the section was too long and time consuming because it required lengthy preparation time. An attempt to shorten the section by simply excluding items did not have the desired effect. Consequently, it was decided to provide the word stems and to use plastic or cardboard letters for the part of the word that needed to be completed. This approach was successful in reducing administration time.

An additional issue that received extensive discussion was the use of one-word stems. This approach was problematic given the difficulty encountered in finding English words that shared a common initial part and contained the required number of blends. Finally, several specific items were also altered to accommodate feedback from the January and March meetings (i.e. avoid the word "gun"; use rime -ad in task # 7 instead of -at).

A Graphophonemic section was added to the Kindergarten Inventory, which combined a phonological awareness task (initial sound identification) with a letter recognition task based on Stuart (1995). This addition provided a model of linking phonological awareness activities to activities related to the acquisition of alphabetic awareness.

Oral reading

Objective: The student is asked to listen to stories or read stories to assess his/her comprehension. If the student is already reading, the text is read aloud by the student. For the student not yet reading, the teacher evaluates comprehension based on a story read to the child.

Revision: Child's oral reading is assessed using a simplified approach for recording miscues (slashing unread or misread words). Teachers then calculate reading accuracy rate for the passage used.

Comprehension

Objective: The ultimate goal of reading is to comprehend what has been read.

Revision: The TPRI provides two passages for listening/reading comprehension at the beginning, middle, and the end of Grades 1 and 2, and three passages at the middle and end of Kindergarten. Questions on narrative passages assess character recall and development, relationship of events, stated facts and details, and inferences. Questions on expository passages in Grades 1 and 2 assess understanding of specific information and concept generalization. Passages have been leveled according to readability and teacher judgment. As in the original TPRI, if additional passages are needed, teachers may select grade-appropriate passages from their school library or basal readers. In doing so, teachers should take care to select passages with which the student has limited or no prior experience.

Passages were selected from different parts of published children's literature by taking into account thematic interest and thematic diversity, as well as book quality, and familiarity. Predictable books were excluded because of text characteristics (inability to construct a wide variety of questions based on the text provided). Efforts were made to ensure that passages formed a complete thematic entity (e.g., containing a beginning, middle, and end; referred to one or more characters; described a setting, situation, or event). After the initial selection, passages were examined for readability by a group of teachers and formal lexile information was obtained. The passages were then assigned grade levels and times of the year such that the order reflected an increasing level of difficulty in text readability. Given the nature of the task at the Kindergarten level (listening as opposed to reading comprehension), some passages may violate this rule. Moreover, in selecting these passages priority was given to the thematic interest of the passage and simple structure instead of text decodeability. Comprehension questions for each passage were initially written to include both literal and implicit components. Passages and questions have undergone several revisions to address issues such as clarity, level of difficulty, and question type. Comprehension questions were scored to reflect the child's ability to address implicit or explicit level questions. Accordingly, the format was modified and questions were re-written to ensure that at least three literal and two implicit questions were included in each passage.

TPRI Field Study

Design

In May, 1998, CARS conducted a pilot study of the TPRI. The field study involved 32 teachers of kindergarten and first grade, 128 kindergarten students, and 144 first grade students in four elementary schools in the Houston Independent School District (HISD). In order to expedite data collection and to compare TPRI assessments with other reading assessments for validity purposes, the field study data were collected as part of a larger study of children supported by the National Institute of Child Health and Human Development (NICHD). In each classroom from 4 schools of the larger study, 8 students were randomly selected from the sample of all NICHD study participants to participate in the field study. This approach had the advantage of working with a group of children viewed as at-risk for reading problems and of different ethnic characteristics than the sample used for the development of the screens. The primary purpose of this study was to estimate the internal consistency of the tasks using classical test theory and generalizability theory, to collect concurrent validity data, and to evaluate teacher responses to the TPRI. The test-retest component yields a lower bound estimate of test-retest reliability and the study was not designed to provide a strong assessment of test-retest reliability.

Procedures

Data were collected on four forms of the TPRI: 1) end of kindergarten, 2) beginning of first grade, 3) end of Grade 1, and 4) beginning of Grade 2. For the beginning Grade 1 and beginning Grade 2 forms, we used children at the end of KG and end of Grade 1 because of the time at which the assessment was conducted. This is a conservative evaluation of the psychometric properties of the TPRI.

Within each form, all students were assessed using both the screen and inventory portions of the TPRI. After administering both portions of the TPRI, teachers completed a TPRI Field Study Questionnaire, which involved questions about all parts of the field study (difficulty of administration of the TPRI, usefulness of the TPRI, and helpfulness of training).

The participating teachers received one day of training from CARS staff. While substitute teachers covered the classrooms, the teachers collected TPRI data on two occasions, with a child they taught and a child they did not teach. In two schools, teachers assessed kindergarten students using Form 1 (end of Kindergarten), and in two other schools using Form 2 (beginning of Grade 1). Similarly, two schools assessed Grade 1 children using Form 3 (end of Grade 1), while teachers in the other two schools used Form 4 (beginning of Grade 2).

Teachers at each grade level and in each school were randomly assigned to a partner teacher. Each teacher in an assigned pair assessed four students from her own classroom and four students from the partner teacher's classroom on each of two occasions. Students were randomly assigned to one of two groups. One group was tested by their own teacher on the first occasion and by the alternate teacher on the second occasion. For the other group, the order of testing was reversed, with the alternate teacher conducting the first assessment, and the student's own

teacher conducting the second assessment. The design of the study is outlined in Table 22 where each group of students represents eight students:

Table 22.
Design of the TPRI Field Study

Group	Schools	Number	Form	May 5 Teacher	May 11 Teacher
K	1-2	56	1	Own	Alternate
K	1-2	56	1	Alternate	Own
K	3-4	55	2	Own	Alternate
K	3-4	55	2	Alternate	Own
1	1-2	44	3	Own	Alternate
1	1-2	44	3	Alternate	Own
1	3-4	62	4	Own	Alternate
1	3-4	62	4	Alternate	Own

Reliability Analyses

The reliability of each of the TPRI tasks was evaluated using both classical test theory and generalizability theory. Classical test theory indices included internal consistency estimates, test-retest reliability estimates, item difficulty parameters, and item-to-total correlations. Cronbach's alpha coefficient ranges from 0 to 1.0 and was reported as the index of internal consistency. High alpha coefficients indicate that item variability is small relative to total test variability, or that all items perform similarly. We evaluated the practical significance of the reliability coefficients as outlined in Table 23. The estimates of practical significance in Table 23 are arbitrary. Because the screen was used to make decisions about individual children, we required coefficients at least in the "good" or "excellent" range. Since there were often only a few items in the inventory tasks, we expected a range of coefficients. Although we required a median in the "good" or better range, we set a lower bound of .40 as "adequate."

Table 23.
Practical Significance of Reliability Estimates by Magnitude

Magnitude of Index	Practical Significance
0-.39	Poor
.40 -.59	Adequate
.60 -.79	Good
.80 - 1.0	Excellent

The correlations between scores at Occasion 1 and Occasion 2 provide estimates of stability over time. Since each student was rated by two different teachers, the test-retest correlation represents a lower-bound estimate of reliability over time. The estimates that are provided will consistently underestimate the true test-retest reliability because different raters were used. Practically, the magnitude of the test-retest correlations was also interpreted similarly to the internal consistency estimates in Table 23.

In Classical Test Theory item difficulty is indexed by the percentage of students who pass an item. Consequently, a difficulty parameter between .7 and 1.0 indicates that an item was not of great difficulty for the students. Practical guidelines for interpreting these percentages are given in Table 24. Item-to-total correlations provided a measure of how students' responses on an item relate to students' total test score. In Classical Test Theory, the item-to-total correlation is a measure of item validity. The practical significance of item-to-total correlation coefficients was interpreted using the criteria given in Table 24.

Table 24.
Predicted Significance of Item Difficulty Indices and Item-Total Correlations by Magnitude

Magnitude of Index	Practical Significance of difficulty index	Practical Significance of item-total correlation
0 - .19	Very Difficult	Poor
.20 - .39	Difficult	Adequate
.40 - .59	Moderately Difficult	Good
.60 - .79	Easy	Excellent
.80 - 1.0	Very Easy	Excellent

In Generalizability Theory, the dependability index expresses the ratio of variability across individuals to variability across individuals, raters, the interaction of raters and individuals, or other sources of random or systematic errors. The dependability index ranges from 0 to 1, and a high index illustrates that task score variation is due to differences between students and not sources of error. Again, interpretation is according to the guidelines in Table 23.

Rater agreement was evaluated at the task level using a difficulty index and Cohen's Kappa. The difficulty index was reported at each occasion and indicated the percentage of students who scored at or above the mastery criterion. Since Cohen's Kappa assesses the degree of agreement across raters when chance agreement is taken into account, it was used to assess teacher agreement on the scoring of the TPRI tasks. The conventional significance of kappa is outlined in Table 25.

Table 25.
Clinical Significance for Kappa Estimates of Inter-Rater Reliability

Kappa Estimates	Clinical Significance
0 - .39	Poor
.40 - .59	Adequate
.60 - .74	Good
.75 – 1.0	Excellent

End of kindergarten analysis

Table 26 summarizes the different indices of reliability for the End of Kindergarten Screen and Kindergarten Inventory. For the screen, all indices of reliability were in the Excellent range: alpha (.92), test-retest (.87), and generalizability (.76).

The inventory also showed reliability, with median values in the excellent to good range: alpha (.89), test-retest (.60), and generalizability (.70) Weak areas largely included Book and Print Awareness. The Comprehension stories varied in their reliability characteristics because of variations in the difficulty levels of the stories. A detailed description of reliability characteristics follows:

Kindergarten Screen

1. **Letter-name Identification** - This 10-item task was administered to 44 kindergarten students on both assessment occasions, 12 only on the first occasion, and 11 only on the second occasion. The distribution of total scores was negatively skewed and demonstrated that 69% of all students scored 10 on the task. Internal consistency estimates were in the excellent range at each occasion ($\alpha=.92$ for both). The test-retest correlation ($r=.95$) demonstrated that scores varied little over the week between occasions and over the two sets of raters. Most items in this task were very easy for the students (% correct ranged from .82 - .91). Item-to-total correlations were in the good range (.59 to .79). Using a mastery criterion of 8 or more correct, 82% of students achieved mastery on Occasion 1 and 80% on Occasion 2. The dependability index for this task indicated that differences between students accounted for most of the variability in scores ($\phi = .75$). Agreement between raters was excellent as evidenced by $\kappa=.93$. Raters scored 35 students above criterion on both occasions, 8 students below criterion on both occasions, and 1 student above criterion the first occasion and below criterion the second occasion.

1. **Letter-sound Identification** - This 10-item task was administered to 56 kindergarten students on both assessment occasions, 8 only on the first occasion, and 7 only on the second occasion. The distribution of total scores was negatively skewed and U-shaped, and demonstrated that 28% of all students scored 10 on the task and 15% scored 0. Internal consistency estimates were in the excellent range at each occasion ($\alpha=.93$ for both). The test-retest correlation ($r=.87$) demonstrated that scores varied little over the week between occasions and over the two sets of raters. Items in this task ranged from difficult to easy for the students (% correct ranged from .34-.70). Item-to-total correlations were in the good to excellent range (.59 to .87). Using a mastery criterion of 8 or more correct, 48% of students achieved mastery on Occasion 1 and 54% on Occasion 2. The high dependability index for this task indicated that differences between students accounted for almost all of the variability in scores ($\phi = .94$). Agreement between raters was good as evidenced by kappa=.61. Raters scored 23 students above criterion on both occasions, 22 students below criterion on both occasions, 4 students above criterion only on the first occasion and below criterion the second occasion, and 7 students below criterion the first occasion and above 8 the second occasion. This task correlated most highly with the Letter-to-Sound Linking B task ($r=.81$), which was not surprising because both of these tasks required students to associate grapheme-phoneme pairs.

2. **Blending** - This 8-item task was administered to 55 kindergarten students on both assessment occasions, 8 only on the first occasion, and 6 only on the second occasion. The distribution of total scores was U-shaped and demonstrated that 28% of all students scored an 8 on the task and 22% scored a 0. Internal consistency was found to be excellent at each occasion ($\alpha=.92$ for Occasion 1 and $\alpha=.94$ for Occasion 2). The test-retest correlation ($r=.66$) demonstrated that scores varied somewhat over the week between occasions and over the two sets of raters. Items in this task ranged from moderately difficult to easy for the students (% correct ranged from .44-.64). Item-to-total correlations were good to excellent (.67 to .83). Using a mastery criterion of 6 or more correct, 47% of students achieved mastery on Occasion 1 and 51% on Occasion 2. The dependability index for this task indicated that differences between students accounted for almost all of the variability in scores ($\phi = .76$). Agreement between raters was good as evidenced by kappa=.64. Raters scored 22 students above criterion and 23 students below criterion on both occasions, 4 students above criterion only on the first occasion, and 6 students above criterion only on the second occasion. This task correlated most highly with the Blending Word Parts and Blending Phonemes tasks ($r=.86$ and $r=.84$, respectively), which was expected since all three tasks assessed blending skills.

Kindergarten Inventory

1. **Book and Print Awareness** - This 5-item task was administered to 56 kindergarten students on both assessment occasions, 8 only on the first occasion, and 6 only on the second occasion. The distribution of total scores was negatively skewed and demonstrated that 85% of all students scored 3-5 on the task. Internal consistency estimates ranged from adequate to poor across the two occasions ($\alpha=.42$ for Occasion 1 and $\alpha=.37$ for Occasion 2). The test-retest correlation ($r=.50$) demonstrated that scores were marginally stable over the week between occasions and over the two sets of raters. Items in this task varied in difficulty for the students (% correct ranged from .30-.98). Item-to-total correlations were in the low

range (.11 to .31). Using a mastery criterion of 3 or more correct, 50% of students achieved mastery on Occasion 1 and 55% on Occasion 2. The dependability index for this task indicated that differences between students accounted for much of the variability in scores ($\phi = .70$). Agreement between raters was adequate as evidenced by $\kappa = .46$. Raters scored 44 students above criterion and, 4 students below criterion on both occasions, 3 students above criterion only on the first occasion, and 4 students above criterion only on the second occasion.

2. **Rhyming** - This 5-item task was administered to 56 kindergarten students on both assessment occasions, 8 only on the first occasion, and 6 only on the second occasion. The distribution of total scores was negatively skewed and U-shaped, and demonstrated that 45% of all students correctly answered all items and 22% incorrectly answered all items on the task. Internal consistency was excellent at each occasion ($\alpha = .89$ for Occasion 1 and $\alpha = .92$ for Occasion 2). The test-retest correlation ($r = .84$) demonstrated that scores varied little over the week between occasions and over the two sets of raters. Items in this task were moderately difficult to easy for the students (% correct ranged from .57 - .68). Item-to-total correlations were in the good to excellent range (.72 to .86). Using a mastery criterion of 4 or more correct, 55% of students achieved mastery on Occasion 1 and 61% on Occasion 2. The dependability index for this task indicated that differences between students accounted for most of the variability in scores ($\phi = .93$). Agreement between raters was good as evidenced by $\kappa = .67$. Raters scored 28 students above criterion and, 19 students below criterion on both occasions, 3 students above criterion only on the first occasion and 6 students above criterion only on the second occasion. This task correlated most highly with the Blending Phonemes Task ($r = .76$).
3. **Blending Word Parts** - This 5-item task was administered to 29 kindergarten students on both assessment occasions, 17 only on the first occasion, and 11 only on the second occasion. The distribution of total scores was negatively skewed and U-shaped, and demonstrated that 49% of all students correctly answered all items and 21% incorrectly answered all items on the task. Internal consistency was excellent at each occasion ($\alpha = .87$ for Occasion 1 and $\alpha = .95$ for Occasion 2). The test-retest correlation ($r = .60$) demonstrated that scores varied some over the week between occasions and over the two sets of raters. Items in this task ranged from moderately difficult to easy for the students (% correct ranged from .55 - .79).

Table 26.
Reliability Information for the TPRI End of Kindergarten Analysis

	Alpha Time 1	Alpha Time 2	*Test-Retest	Generalizability Coefficient	Test Difficulty Time 1	Test Difficulty Time 2	Kappa for Task
SCREEN							
Letter-name	.92	.92	.95	.75	.82	.80	.93
Letter-sound	.93	.93	.87	.94	.48	.54	.61
Blending	.92	.94	.66	.76	.47	.51	.64
INVENTORY							
Book and Print Awareness	.42	.37	.50	.70	.50	.55	.46
Rhyming	.89	.92	.84	.93	.55	.61	.67
Blending Word Parts	.87	.95	.60	.85	.59	.72	.40
**Blending Phonemes	.92	.89	.55	.91	.79	.84	.13
**Deleting Initial Sounds	.92	.94	.63	.73	.50	.63	.75
**Deleting Final Sounds	.90	.94	.72	.70	.46	.38	.84
Letter-name Identification	.97	.97	.92	.96	.80	.81	.82
Letter to Sound Linking A	.83	.88	.54	.85	.77	.84	.49
Letter to Sound Linking B	.81	.89	.51	.83	.77	.86	.55
Listening Comprehension 1	.70	.67	.46	.51	.20	.39	.22
Listening Comprehension 2	.81	.70	.54	.75	.57	.70	.43
Listening Comprehension 3	.72	.65	.63	.70	.47	.64	.61

Note.

* Test –Retest reliability estimate is a lower-bound estimate, since two different raters were used. One week passed between the two occasions.

** Sample sizes for both occasions fell below 20

Item-to-total correlations were in the good to excellent range (.72 to .86). Using a mastery criterion of 4 or more correct, 59% of students achieved mastery on Occasion 1 and 72% on Occasion 2. The high dependability index for this task indicated that differences between students accounted for most of the variability in scores ($\phi = .85$). Agreement between raters on mastery decisions was not as good as expected as evidenced by $\kappa = .40$; however, training of the test administrators should improve rater agreement. Raters scored 15 students above criterion on both occasions, and 6 students below criterion on both occasions. However, 2 students were scored above criterion only on the first occasion, and 6 students were scored above criterion only on the second occasion.

4. **Blending Phonemes** - This 5-item task was administered to 19 kindergarten students on both assessment occasions, 17 only on the first occasion, and 13 only on the second occasion. Because of the small sample size, reliability estimates are not considered stable; however, they are reported as preliminary estimates. The distribution of total scores was negatively skewed and demonstrated that 59% of all students correctly answered all items on the task. Internal consistency was excellent at each occasion ($\alpha = .92$ for Occasion 1 and $\alpha = .89$ for Occasion 2). The test-retest correlation ($r = .55$) demonstrated that scores varied some over the week between occasions and over the two sets of raters. Items in this task were not difficult for the students (% correct ranged from .68 -1.0). Item-to-total correlations were good to excellent (.62 to .87). Using a mastery criterion of 4 or more correct, 79% of students achieved mastery on Occasion 1 and 84% on Occasion 2. The high dependability index for this task indicated that differences between students accounted for most of the variability in scores ($\phi = .91$). Agreement between raters was not high as evidenced by $\kappa = .13$, however this estimate was based on only 19 subjects. A larger sample would provide a better estimate of rater agreement. Raters scored 13 students above criterion and 1 student below criterion on both occasions, 2 students above criterion only on the first occasion and 3 students above criterion only on the second occasion. This task correlated most highly with the Blending Word Parts Task ($r = .85$) which was not surprising since both of these tasks assessed blending skills.
5. **Deleting Initial Sounds** - This 5-item task was administered to 16 kindergarten students on both assessment occasions, 16 only on the first occasion, and 12 only on the second occasion. Because of the small sample size, reliability estimates are not considered stable; however, they are reported as preliminary estimates. The distribution of total scores was U-shaped and demonstrated that 37% of all students correctly answered all items and 32% incorrectly answered all items on the task. Internal consistency was excellent at each occasion ($\alpha = .92$ for Occasion 1 and $\alpha = .94$ for Occasion 2). The test-retest correlation ($r = .63$) demonstrated that scores varied some over the week between occasions and over the two sets of raters. Items in this task ranged from moderately difficult to easy for the students (% correct ranged from .50 - .75). Item-to-total correlations were excellent (.74 to .90). Using a mastery criterion of 4 or more correct, 50% of students achieved mastery on Occasion 1 and 63% on Occasion 2. The dependability index for this task indicated that differences between students accounted for most of the variability in scores ($\phi = .73$). Agreement between raters was good as evidenced by $\kappa = .75$. Raters scored 8 students above criterion on both occasions, 6 students below criterion on both occasions, and 2 students above criterion only on the second occasion. This task correlated most highly with the Letter-sound Task ($r = .71$).

6. **Deleting Final Sounds** - This 5-item task was administered to 13 kindergarten students on both assessment occasions, 12 only on the first occasion, and 10 only on the second occasion. Because of the small sample size, reliability estimates are not considered stable; however, they are reported as preliminary estimates. The distribution of total scores was U-shaped and demonstrated that 46% of all students correctly answered all items and 38% incorrectly answered all items on the task. Internal consistency estimates were excellent at each occasion ($\alpha=.92$ for Occasion 1 and $\alpha=.94$ for Occasion 2). The test-retest correlation was good ($r=.63$) and demonstrated that scores varied some over the week between occasions and over the two sets of raters. Items in this task were moderately difficult to easy for the students (% correct ranged from .50 -.75). Item-to-total correlations were in the good to excellent range (.75 to .90). Using a mastery criterion of 4 or more correct, 50% of students achieved mastery on Occasion 1 and 72% on Occasion 2, possibly reflecting repeated exposure. The high dependability index for this task indicated that differences between students accounted for most of the variability in scores ($\phi = .85$). Agreement between raters was estimated on only 13 subjects and was not as high as expected as evidenced by $\kappa=.40$. The Occasion 1 by Occasion 2 plot of student scores revealed that some teachers scored their own students higher than students from other classes. Raters scored 5 students above criterion on both occasions, 7 students below criterion on both occasions, and 1 student above criterion only on the first occasion.
7. **Letter-name Identification** - This 26-item task was administered to 54 kindergarten students on both assessment occasions, 8 only on the first occasion, and 8 only on the second occasion. The distribution of total scores was negatively skewed and demonstrated that 65% of all students correctly answered all items on the task. Internal consistency was excellent at each occasion ($\alpha=.97$ on both occasions). The test-retest correlation ($r=.92$) demonstrated that scores varied little over the week between occasions and over the two sets of raters. Items in this task were easy to very easy for the students (% correct ranged from .76 -.96). Item-to-total correlations were in the adequate to excellent range (.56 to .87). Using a mastery criterion of 20 or more correct, 80% of students achieved mastery on Occasion 1 and 81% on Occasion 2. The dependability index for this task was excellent and indicated that differences between students accounted for almost all of the variability in scores ($\phi = .96$). Agreement between raters was also excellent as indicated by κ of .82. Raters scored 42 students above criterion and 9 students below criterion on both occasions. In contrast, 1 student scored above criterion only on the first occasion and 2 students were scored above criterion only on the second occasion. This task correlated most highly with the Letter-to-Sound Linking B Task ($r=.93$) which was not surprising since both of these tasks focused on single graphemes.
8. **Letter to Sound Linking A** - This 5-item task required students to identify the first sound in a word. The task was administered to 43 kindergarten students on both assessment occasions, 12 only on the first occasion, and 12 only on the second occasion. The distribution of total scores was negatively skewed and demonstrated that 56% of all students correctly answered all items on the task. Internal consistency was excellent at each occasion ($\alpha=.83$ for Occasion 1 and $\alpha=.88$ for Occasion 2). The test-retest correlation ($r=.54$) demonstrated that scores varied some over the week between occasions and over the two sets

of raters. Items in this task were easy to very easy for the students (% correct ranged from .77 -.91). Item-to-total correlations ranged from adequate to excellent (.47 to .82). Using a mastery criterion of 4 or more correct, 77% of students achieved mastery on Occasion 1 and 84% on Occasion 2. The high dependability index for this task indicated that differences between students accounted for most of the variability in scores ($\phi = .85$). Agreement between raters was fair as evidenced by $\kappa = .49$. Raters scored 31 students above criterion on both occasions, 5 students below criterion on both occasions, 2 students above criterion only on the first occasion and 5 above criterion only on the second occasion.

9. **Letter to Sound Linking B** - This 5-item task required students to select the correct letter from 3 choices that represent the target sound. The task was administered to 43 kindergarten students on both assessment occasions, 13 only on the first occasion, and 12 only on the second occasion. The distribution of total scores was negatively skewed and demonstrated that 62% of students correctly answered all items on the task. Internal consistency was excellent at each occasion ($\alpha = .81$ for Occasion 1 and $\alpha = .89$ for Occasion 2). The test-retest correlation ($r = .51$) demonstrated that scores varied over the week between occasions and over the two sets of raters. Items in this task were easy to very easy for the students (% correct ranged from .72 -.91). Item-to-total correlations ranged from adequate to excellent (.42 to .89). Using a mastery criterion of 4 or more correct, 77% of students achieved mastery on Occasion 1 and 86% on Occasion 2, possibly reflecting exposure over a short time period. The dependability index for this task indicated that differences between students accounted for most of the variability in scores ($\phi = .83$). Agreement between raters was adequate as evidenced by $\kappa = .55$. Raters scored 32 students above criterion and 5 students below criterion on both occasions, 1 student above criterion only on the first occasion and 5 students above criterion only on the second occasion.

10. **Listening Comprehension 1** - This 5-item task was administered to 55 kindergarten students on both assessment occasions, 7 only on the first occasion, and 7 only on the second occasion. The distribution of total scores approximated a normal distribution. Internal consistency was good at each occasion ($\alpha = .70$ for Occasion 1 and $\alpha = .67$ for Occasion 2). The test-retest correlation ($r = .46$) demonstrated that scores varied some over the week between occasions and over the two sets of raters. Items in this task ranged from difficult to easy for the students (% correct ranged from .27 -.75). Item-to-total correlations fell in the poor and adequate ranges (.24 to .51). Using a mastery criterion of 4 or more correct, 20% of students achieved mastery on Occasion 1 and 39% on Occasion 2, possibly reflecting repeated exposure. The dependability index for this task was moderate, but indicated that differences between students accounted for much of the variability in scores ($\phi = .51$). Agreement between raters was not high as evidenced by $\kappa = .22$. This relatively poor κ indicates a need for improved scoring criteria for items, increasing the number of questions, and possibly reducing mastery judgements unless more than one story is presented to the child. Raters scored 7 students above criterion on both occasions and 30 students below criterion on both occasions. At the same time, 4 students were scored above criterion only on the first occasion, and 14 students were scored above criterion only on the second occasion.

- 11. Listening Comprehension 2** - This 5-item task was administered to 56 kindergarten students on both assessment occasions, 7 only on the first occasion, and 6 only on the second occasion. The distribution of total scores was negatively skewed and demonstrated that 46% of students correctly answered all items on the task. Internal consistency estimates ranged from good to excellent at each occasion ($\alpha=.81$ for Occasion 1 and $\alpha=.70$ for Occasion 2). The test-retest correlation ($r=.54$) demonstrated that scores varied some over the week between occasions and over the two sets of raters. Items in this task ranged from moderately difficult to very easy for the students (% correct ranged from .50 -.96). Item-to-total correlations ranged from adequate to good (.44 to .61). Using a mastery criterion of 4 or more correct, 57% of students achieved mastery on Occasion 1 and 70% on Occasion 2, possibly reflecting repeated exposure. The dependability index for this task indicated that differences between students accounted for most of the variability in scores ($\phi = .75$). Agreement between raters was adequate as evidenced by kappa=.43. This value for kappa indicates that improved scoring criteria for individual questions or alternatively increasing the number of questions per story may be warranted. Another possibility is to reserve mastery judgements for performance based on both stories, rather than rate stories individually for mastery. Raters scored 28 students above criterion on both occasions, 13 students below criterion on both occasions, 4 students above criterion only on the first occasion and 11 students above criterion only on the second occasion.
- 12. Listening Comprehension 3** - This 5-item task was administered to 55 kindergarten students on both assessment occasions, 7 only on the first occasion, and 7 only on the second occasion. The distribution of total scores was negatively skewed and demonstrated that 58% of students scored a 4 or 5 on the task. Internal consistency estimates were good at each occasion ($\alpha=.72$ for Occasion 1 and $\alpha=.65$ for Occasion 2). The test-retest correlation ($r=.63$) demonstrated that scores varied some over the week between occasions and over the two sets of raters. Items in this task ranged from difficult to very easy for the students (% correct ranged from .38 -.93). Item-to-total correlations ranged from poor to adequate (.30 to .57). Using a mastery criterion of 4 or more correct, 47% of students achieved mastery on Occasion 1 and 64% on Occasion 2. The dependability index for this task indicated that differences between students accounted for most of the variability in scores ($\phi = .70$). Agreement between raters was good as evidenced by kappa=.61. Raters scored 25 students above criterion on both occasions, 19 students below criterion on both occasions, 1 student above criterion on the first occasion but not the second, and 10 students above criterion on the second occasion but not on the first.

Beginning Grade 1 analysis

Table 27 provides the reliability estimates for the Beginning Grade 1 screen and Grade 1 inventory. The screen has excellent reliability characteristics (median alpha = .86; median generalizability = .86) with some of the lower-bound test-retest coefficients in the upper part of the good range (median = .76). For the inventory, median alpha (.79) and generalizability (.68) coefficients are in the good range, with adequate test-retest reliability (median = .54).

Book and Print Awareness and Final Consonant Substitution were weak. For the Comprehension 1 story, alpha is poor, but the Generalizability Coefficient is good. A detailed discussion of reliability of the Grade 1 forms follows.

Reliability of Beginning Grade 1 Screen

1. **Letter-Name Identification** - This 10-item task was administered to 47 kindergarten students on both assessment occasions, 15 only on the first occasion, and 9 only on the second occasion. The distribution of total scores was negatively skewed and demonstrated that 69% of all students scored 10 on the task. Internal consistency estimates were good at each occasion ($\alpha=.74$ for Occasion 1 and $\alpha=.80$ for Occasion 2). The test-retest correlation ($r=.80$) demonstrated that scores did not vary much over the week between occasions and over the two sets of raters. Most items in this task were easy to very easy for the students (% correct ranged from .85 - 1.0). Item-to-total correlations ranged from poor to excellent (.04 to .85). Using a mastery criterion of 8 or more correct, 94% of students achieved mastery on Occasion 1 and 89% on Occasion 2, which is not a large difference. The dependability index for this task indicated that differences between students accounted for almost all of the variability in scores ($\phi = .96$). Agreement between raters was good as evidenced by kappa=.73. Raters scored 42 students above criterion on both occasions and 3 students below criterion on both occasions. Two students were scored above criterion on the first occasion but not on the second occasion.
2. **Letter-sound Identification** - This 10-item task was administered to 55 kindergarten students on both assessment occasions, 8 only on the first occasion, and 9 only on the second occasion. The distribution of total scores was negatively skewed and demonstrated that 20% of all students scored 10 on the task. Internal consistency was excellent at each occasion ($\alpha=.86$ for Occasion 1 and $\alpha=.89$ for Occasion 2). The test-retest correlation ($r=.76$) demonstrated that scores did not vary much over the week between occasions and over the two sets of raters. Most items in this task ranged from moderately difficult to easy for the students (% correct ranged from .41 - .76). Item-to-total correlations ranged from poor to good (.38 to .70). Using a mastery criterion of 8 or more correct, 35% of students achieved mastery on Occasion 1 and 51% on Occasion 2. The dependability index for this task indicated that differences between students accounted for most of the variability in scores ($\phi = .76$). Agreement between raters was good as evidenced by kappa=.60. Raters scored 18 students above criterion on both occasions, 26 students below criterion on both occasions, 1 student above criterion only on the first occasion and 10 students above criterion only on the second occasion. This task correlated most highly with the Initial Consonant Substitution (ICS) task ($r=.69$). This high degree of relation to ICS could be explained by the

composition of the ICS Task, for which 4 of the 5 initial consonants used were of the type that the letter-sound is incorporated in the letter-name (e.g. mad, dad).

3. **Word Reading** - This 10-item task was administered to 44 kindergarten students on both assessment occasions, 18 only on the first occasion, and 9 only on the second occasion. The distribution of total scores was positively skewed and demonstrated that 83% of all students scored 0 on the task. Internal consistency was excellent at each occasion ($\alpha=.91$ for Occasion 1 and $\alpha=.89$ for Occasion 2). The test-retest correlation ($r=.87$) demonstrated that scores varied little over the week between occasions and over the two sets of raters. Most items in this task were very difficult for the students (% correct ranged from .02 - .14). Item-to-total correlations ranged from adequate to excellent (.46 to .84). Using a mastery criterion of 8 or more correct, 2% of students achieved mastery on Occasion 1 and 2% on Occasion 2. The high dependability index for this task was only estimable using a technique which assumes normality of scores; therefore the index is only a rough estimate ($\phi = .90$), but indicates that differences between students account for most of the variability in scores. Agreement between raters was excellent as evidenced by $\kappa=1.0$. Raters scored 1 student above criterion on both occasions and 43 students below criterion on both occasions. No students were score above criterion on only one of the two occasions.

4. **Blending Phonemes** - This 6-item task was administered to 54 kindergarten students on both assessment occasions, 10 only on the first occasion, and 8 only on the second occasion. The distribution of total scores was positively skewed and demonstrated that 24% of all students scored 0 on the task. Internal consistency was excellent at each occasion ($\alpha=.80$ for Occasion 1 and $\alpha=.89$ for Occasion 2). The test-retest correlation ($r=.70$) demonstrated that scores did not vary much over the week between occasions and over the two sets of raters. Most items in this task ranged from difficult to easy for the students (% correct ranged from .20 - .60). Item-to-total correlations ranged from poor to excellent (.39 to .73), with most in the good to excellent range. Using a mastery criterion of 5 or more correct, 20% of students achieved mastery on Occasion 1 and 43% on Occasion 2, possibly reflecting repeated exposure over a short time interval. The high dependability index for this task indicated that differences between students accounted for most of the variability in scores ($\phi = .85$). Agreement between raters was not as good as expected as evidenced by $\kappa=.35$; however, agreement should improve with additional training. Raters scored 9 students above criterion on both occasions, 29 students below criterion on both occasions, 2 student above criterion only on the first occasion and below 5 the second occasion, and 14 students above criterion only on the second occasion. This task correlated most highly with the Letter-sound Task ($r=.69$).

Table 27
Reliability Information for the TPRI Beginning of First Grade Form

	Alpha Time 1	Alpha Time 2	*Test-Retest	Generalizability Coefficient	Test Difficulty Time 1	Test Difficulty Time 2	Kappa for Task
SCREEN							
Letter-name	.74	.80	.80	.96	.94	.89	.73
Letter-sound	.86	.89	.76	.86	.35	.51	.60
Word Reading	.91	.89	.87	.90	.02	.02	1.0
Blending Phonemes	.80	.89	.70	.85	.20	.43	.35
INVENTORY							
Book and Print Awareness	.44	.17	.20	.48	.75	.79	.01
Rhyming	.85	.92	.78	.92	.73	.73	.54
Blending Onset Rimes	.87	.83	.64	.84	.55	.71	.56
Blending Phonemes	.88	.74	.54	.87	.82	.86	.49
Deleting Initial Sounds	.92	.94	.39	.74	.55	.70	.48
**Deleting Final Sounds	.94	.85	.21	.42	.75	.58	-.09
Initial Consonant Substitution	.91	.79	.84	.90	.75	.88	.60
Final Consonant Substitution	.59	.62	.12	.63	.68	.80	.25
Medial Vowel Substitution	.63	.79	.54	.60	.39	.54	.30
**Initial Blend Substitution	.88	.67	-.13	.36	.55	.55	.27
**Final Blend Substitution	.94	.79	-.33	.61	.67	1.0	0
Comprehension 1	.36	.29	.59	.76	.71	.89	.35
Comprehension 2	.66	.71	.62	.68	.20	.34	.47

Note.

* Test –Retest reliability estimate is a lower-bound estimate, since two different raters were used. One week passed between the two occasions.

** Sample sizes for both occasions fell below 20.

Reliability of Beginning First Grade Inventory

1. **Book and Print Awareness** - This 5-item task was administered to 55 kindergarten students on both assessment occasions, 9 only on the first occasion, and 8 only on the second occasion. The distribution of total scores was negatively skewed and demonstrated that 77% of all students scored 4-5 on the task. Internal consistency ranged from poor to adequate across the two occasion ($\alpha=.44$ for Occasion 1 and $\alpha=.17$ for Occasion 2). The poor test-retest correlation ($r=.20$) demonstrated that scores varied over the week between occasions and over the two sets of raters. This variability indicates the need to improve administration and scoring criteria for this task. For example, item 2 requires students to point to a sentence. The potential for inconsistent scoring is high in tasks such as this. Most items in this task ranged from difficult to very easy for the students (% correct ranged from .35 - 1.0). Item-to-total correlations ranged from poor to adequate (.05 to .39). Using a mastery criterion of 4 or more correct, 75% of students achieved mastery on Occasion 1 and 79% on Occasion 2. The dependability index for this task indicated that differences between students did not account for most of the variability in scores ($\phi = .48$). Agreement between raters was poor evidenced by $\kappa=.01$. Raters scored 48 students above criterion on both occasions and no students below criterion on both occasions. At the same time, 3 students were scored above criterion only on the first occasion, and 4 students were scored above criterion only on the second occasion. This task did not correlate above $r = .50$ with any other task.

2. **Rhyming** - This 5-item task was administered to 55 kindergarten students on both assessment occasions, 9 only on the first occasion, and 8 only on the second occasion. The distribution of total scores was negatively skewed and demonstrated that 77% of all students scored 4-5 on the task. Internal consistency estimates were excellent at each occasion ($\alpha=.85$ for Occasion 1 and $\alpha=.92$ for Occasion 2). The test-retest correlation ($r=.78$) demonstrated that scores did not vary much over the week between occasions and over the two sets of raters. Most items in this task ranged from easy to very easy for the students (% correct ranged from .71 - .82). Item-to-total correlations ranged from good to excellent (.59 to .88). Using a mastery criterion of 4 or more correct, 73% of students achieved mastery on both occasions. The high dependability index for this task indicated that differences between students accounted for almost all of the variability in scores ($\phi = .92$). Agreement between raters was adequate as evidenced by $\kappa=.54$. Raters scored 35 students above criterion on both occasions, 10 students below criterion on both occasions, 5 students above criterion on the first occasion but below criterion on the second occasion, and 5 students were scored with the opposite pattern. This task correlated most highly with the Final Consonant Substitution Task ($r=.46$) which is not surprising because both of these tasks require that the student focus on the endings of words.

3. **Blending Onset-Rime** - This 5-item task was administered to 38 kindergarten students on both assessment occasions, 31 only on the first occasion, and 11 only on the second occasion. The distribution of total scores was negatively skewed and demonstrated that 38% of all students correctly answered all items on the task. Internal consistency estimates were excellent at each occasion ($\alpha=.82$ for Occasion 1 and $\alpha=.83$ for Occasion 2). The test-retest correlation ($r=.64$) demonstrated that scores varied some over the week between occasions

and over the two sets of raters. Items in this task ranged from moderately difficult to very easy for the students (% correct ranged from .50 -.87). Item-to-total correlations ranged from adequate to excellent (.45 to .79). Using a mastery criterion of 4 or more correct, 55% of students achieved mastery on Occasion 1 and 71% on Occasion 2. The dependability index for this task indicated that differences between students accounted for most of the variability in scores ($\phi = .84$). Agreement between raters was adequate as evidenced by $\kappa = .56$. Raters scored 20 students above criterion on both occasions, 10 students below criterion on both occasions, 1 student above criterion only on the first occasion, and 7 students above criterion only on the second occasion. This task correlated most highly with the Blending Phonemes Task ($r = .69$).

4. **Blending Phonemes** - This 5-item task was administered to 22 kindergarten students on both assessment occasions, 14 only on the first occasion, and 16 only on the second occasion. The distribution of total scores was negatively skewed and demonstrated that 76% of all students scored 4-5 on the task. Internal consistency was good to excellent at each occasion ($\alpha = .88$ for Occasion 1 and $\alpha = .74$ for Occasion 2). The test-retest correlation ($r = .54$) demonstrated that scores varied some over the week between occasions and over the two sets of raters. Items in this task ranged from moderately easy to very easy for the students (% correct ranged from .61 -.95). Item-to-total correlations ranged from good to excellent (.30 to .91). Using a mastery criterion of 4 or more correct, 82% of students achieved mastery on Occasion 1 and 86% on Occasion 2. The dependability index for this task indicated that differences between students accounted for most of the variability in scores ($\phi = .87$). Agreement between raters was adequate as evidenced by $\kappa = .49$. Raters scored 17 students above criterion and 2 students below criterion on both occasions. Raters disagreed on 3 students, scoring 1 above criterion only on the first occasion and 2 students above criterion only on the second occasion.

5. **Deleting Initial Sounds** - This 5-item task was administered to 20 kindergarten students on both assessment occasions, 13 only on the first occasion, and 14 only on the second occasion. The distribution of total scores was U-shaped and demonstrated that 31% of all students scored 0 and 40% scored a perfect 5 on the task. Internal consistency estimates were excellent at each occasion ($\alpha = .92$ for Occasion 1 and $\alpha = .94$ for Occasion 2). The test-retest correlation ($r = .39$) demonstrated that scores varied over the week between occasions and over the two sets of raters. Items in this task were easy for the students (% correct ranged from .70 -.80). Item-to-total correlations ranged from excellent to outstanding (.72 to .89). Using a mastery criterion of 4 or more correct, 55% of students achieved mastery on Occasion 1 and 70% on Occasion 2. The gain in percent mastery may indicate a practice effect. The dependability index for this task indicated that differences between students accounted for most of the variability in scores ($\phi = .74$). Agreement between raters was adequate as evidenced by $\kappa = .48$. Raters scored 10 students above criterion on both occasions, 5 students below criterion on both occasions, 1 student above criterion on the first occasion only, and 4 students above criterion on the second occasion only. This task correlated most highly with the Blending Onset-Rime Task ($r = .52$) which is not unsurprising since both of these tasks require manipulation of single phonemes.

6. **Deleting Final Sounds** - This 5-item task was administered to 12 kindergarten students on both assessment occasions, 15 only on the first occasion, and 14 only on the second occasion. Because of the small sample size, reliability estimates are not considered stable; however, they are reported as preliminary estimates. The distribution of total scores was U-shaped and demonstrated that 32% of all students scored 0 and 28% scored a perfect 5 on the task. Internal consistency was excellent at each occasion ($\alpha=.94$ for Occasion 1 and $\alpha=.85$ for Occasion 2). The test-retest correlation ($r=.21$) demonstrated that scores varied considerably over the week between occasions and over the two sets of raters. Items in this task ranged from moderately difficult to very easy for the students (% correct ranged from .42 -.92). Item-to-total correlations ranged from adequate to excellent (.49 to .89). Using a mastery criterion of 4 or more correct, 75% of students achieved mastery on Occasion 1 and 58% on Occasion 2. The dependability index for this task indicated that differences between students accounted for some of the variability in scores ($\phi = .42$). Agreement between raters was poor as evidenced by kappa= -.09; however, this estimate is only based on the scores of 12 students. Raters agreed on the scoring of 6 students, with 5 of these scoring above criterion and one scoring below criterion. Of the remaining 6 students, 4 scored above criterion on the first occasion and 2 score above criterion on the second occasion. This task correlated most highly with the Deleting Initial Sounds Task ($r=.62$).
7. **Initial Consonant Substitution** - This 5-item task was administered to 48 kindergarten students on both assessment occasions, 12 only on the first occasion, and 12 only on the second occasion. The distribution of total scores was negatively skewed and demonstrated that 69% of all students scored a perfect 5 on the task. Internal consistency was good to excellent at each occasion ($\alpha=.91$ for Occasion 1 and $\alpha=.79$ for Occasion 2). The test-retest correlation ($r=.84$) demonstrated that scores varied little over the week between occasions and over the two sets of raters. Items in this task were very easy for the students (% correct ranged from .83 -.91). Item-to-total correlations ranged from adequate to excellent (.47 to .88). Using a mastery criterion of 4 or more correct, 75% of students achieved mastery on Occasion 1 and 88% on Occasion 2. The excellent dependability index for this task indicated that differences between students accounted for most of the variability in scores ($\phi = .84$). Agreement between raters was good as evidenced by kappa= .60. Raters scored 38 students above criterion on both occasions, 6 students below criterion on both occasions, and 4 students above criterion only on the second occasion. This task correlated most highly with both Letter-Names and Letter-Sounds Tasks ($r=.69$).
8. **Final Consonant Substitution** - This 5-item task was administered to 12 kindergarten students on both assessment occasions, 15 only on the first occasion, and 14 only on the second occasion. Because of the small sample size, reliability estimates are not considered stable; however, they are reported as preliminary estimates. The distribution of total scores was U-shaped and demonstrated that 32% of all students scored 0 and 28% scored a perfect 5 on the task. Internal consistency was excellent at each occasion ($\alpha=.94$ for Occasion 1 and $\alpha=.85$ for Occasion 2). The test-retest correlation ($r=.21$) demonstrated that scores varied quite a bit over the week between occasions and over the two sets of raters. Items in this task ranged from moderately difficult to very easy for the students (% correct ranged from .42 -.92). Item-to-total correlations ranged from adequate to excellent (.49 to .89). Using a

mastery criterion of 4 or more correct, 75% of students achieved mastery on Occasion 1 and 58% on Occasion 2. The dependability index for this task indicated that differences between students accounted for some of the variability in scores ($\phi = .42$). Agreement between raters was poor as evidenced by kappa = $-.09$. Raters scored 5 students above criterion on both occasions, and 1 student below criterion on both occasions. Raters disagreed on the other 6 students with 4 scoring above criterion on the first occasion and 2 scoring above criterion on the second occasion. This task correlated most highly with the Deleting Initial Sounds Task ($r = .62$).

9. **Medial Vowel Substitution** - This 5-item task was administered to 26 kindergarten students on both assessment occasions, 17 only on the first occasion, and 16 only on the second occasion. The distribution of total scores was slightly negatively skewed and demonstrated that the most common score was 2. Internal consistency was good at each occasion ($\alpha = .63$ for Occasion 1 and $\alpha = .79$ for Occasion 2). The test-retest correlation ($r = .54$) demonstrated that scores varied some over the week between occasions and over the two sets of raters. Items in this task ranged from difficult to easy for the students (% correct ranged from $.38 - .79$). Item-to-total correlations ranged from poor to adequate ($.24$ to $.59$). Using a mastery criterion of 4 or more correct, 39% of students achieved mastery on Occasion 1 and 54% on Occasion 2. The dependability index for this task indicated that differences between students accounted for much of the variability in scores ($\phi = .60$). Agreement between raters was poor as evidenced by kappa = $.30$; however agreement should improve with further training on this task. Raters scored 7 students above 4 on both occasions, 9 students below criterion on both occasions, 3 students above criterion only on the first occasion, and 7 students above criterion only on the second occasion. This task correlated most highly with the Final Blend Substitution Task ($r = .79$).

10. **Initial Blend Substitution** - This 5-item task was administered to 11 kindergarten students on both assessment occasions, 15 only on the first occasion, and 11 only on the second occasion. Because of the small sample size, reliability estimates are not considered stable; however, they are reported as preliminary estimates. The distribution of total scores was U-shaped. Internal consistency ranged from good to excellent across the two occasions ($\alpha = .88$ for Occasion 1 and $\alpha = .67$ for Occasion 2). The test-retest correlation ($r = -.13$) was calculated based only on 11 students' scores and demonstrated that scores varied considerably over the week between occasions and over the two sets of raters. Items in this task ranged from difficult to very easy for the students (% correct ranged from $.36 - .82$). Item-to-total correlations ranged from poor to excellent ($.25$ to $.86$). Using a mastery criterion of 4 or more correct, 55% of students achieved mastery on Occasion 1 and 55% on Occasion 2. The dependability index for this task indicated that differences between students only accounted for some of the variability in scores ($\phi = .36$). Agreement between raters was poor as evidenced by kappa = $.27$; but should improve with further training on this task. Raters agreed on 7 of 11 students, with 4 scoring above criterion and 3 scoring below criterion. The 4 disagreements were evenly divided with 2 students scoring above criterion on the first occasion and 2 scoring above criterion on the second occasion. This task correlated most highly with the Medial Vowel Substitution Task ($r = .59$).

11. **Final Blend Substitution** - This 5-item task was administered to 6 kindergarten students on both assessment occasions, 14 only on the first occasion, and 12 only on the second occasion. Because of the small sample size, reliability estimates are not considered stable; however, they are reported as preliminary estimates. The distribution of total scores was U-shaped and demonstrated that 55% of all students correctly answered all items and 18% incorrectly answered all items on the task. Internal consistency ranged from excellent to good across the 2 occasions ($\alpha=.94$ for Occasion 1 and $\alpha=.79$ for Occasion 2). The test-retest correlation ($r=-.33$) was calculated based only on 6 students and demonstrated that these 6 pairs of scores varied considerably over the week between occasions and over the two sets of raters. Items in this task ranged from easy to very easy for the students (% correct ranged from .67 - 1.0). Item-to-total correlations ranged from poor to excellent (.36 to .93). Using a mastery criterion of 4 or more correct, 67% of students achieved mastery on Occasion 1 and 100% on Occasion 2, possibly reflecting repeated exposure or practice. The dependability index for this task was only estimable using a technique which assumes normality of scores, therefore the index is only a rough estimate ($\phi = .61$). No agreement was found between raters when chance agreement was removed. Raters scored 4 students above criterion on both occasions, and 2 students above criterion only on the second occasion.
12. **Comprehension 1** - This 5-item task was administered to 56 kindergarten students on both assessment occasions, 8 only on the first occasion, and 8 only on the second occasion. The distribution of total scores was negatively skewed and demonstrated that 81% of all students scored 4-5 on the task. Internal consistency estimates at each occasion were not as high as expected ($\alpha=.36$ for Occasion 1 and $\alpha=.29$ for Occasion 2). The low internal consistency estimates may reflect that some questions were implicit and others explicit. The test-retest correlation ($r=.59$) demonstrated that scores varied some over the week between occasions and over the two sets of raters. Items in this task ranged from moderately difficult to very easy for the students (% correct ranged from .46 - 1.0). Item-to-total correlations fell in the poor and adequate ranges (.11 to .34). Using a mastery criterion of 4 or more correct, 71% of students achieved mastery on Occasion 1 and 89% on Occasion 2, possibly reflecting repeated exposure. The dependability index for this task indicated that differences between students accounted for most of the variability in scores ($\phi = .76$). Agreement between raters was not high as evidenced by $\kappa=.35$. The level of agreement suggests the need for more explicit scoring criteria and/or additional training to reduce the effects of subjectivity from raters. Raters scored 39 students above criterion on both occasions and 5 students below criterion on both occasions, but disagreed on the scoring of 12 students with 11 of those scoring above criterion only on the second occasion. This task did not correlate above .40 with any other task, but the task it correlated most highly with was the Book and Print Awareness Task ($r=.36$).
13. **Comprehension 2** - This 5-item task was administered to 56 kindergarten students on both assessment occasions, 8 only on the first occasion, and 8 only on the second occasion. The distribution of total scores was slightly positively skewed and demonstrated that the most common score was a 2. Internal consistency was good at each occasion ($\alpha=.66$ for Occasion 1 and $\alpha=.71$ for Occasion 2). The test-retest correlation ($r=.62$) demonstrated that scores varied some over the week between occasions and over the two sets of raters. Items in this

task ranged from difficult to easy for the students (% correct ranged from .29 - .66). Item-to-total correlations fell in the poor and adequate ranges (.34 to .47). Using a mastery criterion of 4 or more correct, 20% of students achieved mastery on Occasion 1 and 34% on Occasion 2. The moderate dependability index for this task indicated that differences between students accounted for most of the variability in scores ($\phi = .68$). Agreement between raters was adequate as evidenced by kappa=.61. Raters scored 9 students above criterion on both occasions and 35 students below criterion on both occasions, but disagreed on the scoring of 12 students with 10 scoring above criterion only on the second occasion. This task correlated most highly with the first comprehension task ($r=.38$).

End of grade 1 analysis

Table 28 summarizes the reliability data for the End of Grade 1 screen and the Grade 1 inventory administered at the end of Grade 1. Reliability estimates were excellent to good for the Word Reading screening task and good for the Blending Phonemes screening task. Reliabilities for the inventory tasks are good for alpha (median = .66), excellent for generalizability (median = .81), and adequate for test-retest (median = .42). Reliability is again poor for Book and Print Awareness. Internal consistency estimates for the first occasion and generalizability coefficients are generally in the good to excellent range except for Blending Phonemes and Comprehension 1. More variability is apparent in the estimates from the second measurement occasion. A detailed discussion of each task follows.

Reliability of End of Grade 1 Screen

1. **Word Reading Task** - This 10-item task was administered to 44 first grade students on both assessment occasions, 4 only on the first occasion, and 4 only on the second occasion. The distribution of total scores was positively skewed and demonstrated that 29% of all students scored 0 on the task. Internal consistency was excellent at each occasion ($\alpha=.88$ for Occasion 1 and $\alpha=.89$ for Occasion 2). The test-retest correlation ($r=.93$) demonstrated that scores varied little over the week between occasions and over the two sets of raters. Most items in this task ranged from very difficult to moderately difficult for the students (% correct ranged from .14 - .50). Item-to-total correlations ranged from adequate to good (.43 to .77). Using a mastery criterion of 8 or more correct, 9% of students achieved mastery on Occasion 1 and 16% on Occasion 2. The dependability index for this task indicated that differences between students accounted for most of the variability in scores ($\phi = .95$). Agreement between raters was good as evidenced by kappa=.69. Raters agreed on the scoring of 40 students, with 37 of those scoring below criterion. Disagreements were found on only 3 students, all of whom scored above criterion only on the second occasion. This task correlated most highly with the Deleting Final Sounds Task ($r=.57$).
2. **Blending Phonemes** - This 6-item task was administered to 44 first grade students on both assessment occasions, 4 only on the first occasion, and 4 only on the second occasion. The distribution of total scores approximated a normal distribution. Internal consistency was estimated to be good at each occasion ($\alpha=.74$ for Occasion 1 and $\alpha=.75$ for Occasion 2). The test-retest correlation ($r=.51$) demonstrated that scores varied some over the week between occasions and over the two sets of raters. Most items in this task ranged from difficult to

very easy for the students (% correct ranged from 25 - .80). Item-to-total correlations ranged from poor to good (.37 to .62). Using a mastery criterion of 5 or more correct, 16% of students achieved mastery on Occasion 1 and 20% on Occasion 2. The dependability index for this task indicated that differences between students accounted for much of the variability in scores ($\phi = .58$). Agreement between raters was poor as evidenced by $\kappa = .09$. Raters agreed on 30 students, 27 of which were below criterion, but disagreed on 14 students, of which 6 scored above criterion only on the first occasion and 8 scored above criterion only on the second occasion. This task correlated most highly with the Word Reading Task ($r = .56$).

Reliability of End Grade 1 Inventory

1. **Book and Print Awareness** - This 5-item task was administered to 44 first grade students on both assessment occasions, 4 only on the first occasion, and 4 only on the second occasion. The distribution of total scores demonstrated that 100% of students scored 4-5 on the task. Because all students scored a perfect 5 on 4 of the 5 items at Occasion 1, internal consistency was not estimable. At Occasion 2 all students scored a perfect 5 on 3 of the 5 items. Most items in this task were very easy for the students (% correct ranged from .77 - 1.0). Item-to-total correlations were not estimable since most items were correctly answered by all students. Using a mastery criterion of 4 or more correct, 100% of students achieved mastery on both occasions. Agreement between raters was perfect, since all students mastered the task on both occasions. This task correlated most highly with the Blending Phonemes Task ($r = .56$).
2. **Rhyming** - - This 5-item task was administered to 44 first grade students on both assessment occasions, 4 only on the first occasion and 4 only on the second occasion. The distribution of total scores was negatively skewed and demonstrated that 76% of all students correctly answered all items on the task. Internal consistency was excellent at each occasion ($\alpha = .83$ for Occasion 1 and $\alpha = .80$ for Occasion 2). The test-retest correlation ($r = .75$) demonstrated that scores did not vary much over the week between occasions and over the two sets of raters. Most items in this task were very easy for the students (% correct ranged from .82 - 95). Item-to-total correlations ranged from adequate to excellent (.30 to .77). Using a mastery criterion of 4 or more correct, 80% of students achieved mastery on Occasion 1 and 93% on Occasion 2. The high dependability index for this task indicated that differences between students accounted for almost all of the variability in scores ($\phi = .98$). Nonetheless, chance adjusted agreement between raters was poor on mastery decisions as evidenced by $\kappa = .26$. Raters agreed on 36 students, with 34 scoring above criterion. Disagreements were found for 8 students, most of whom (7) were scored above criterion only on the second occasion. This task correlated most highly with the Blending Onset-Rime Task ($r = .69$).
3. **Blending Onset-Rime** - This 5-item task was administered to 36 first grade students on both assessment occasions, 5 only on the first occasion, and 9 only on the second occasion. The distribution of total scores was negatively skewed and demonstrated that 91% of all students correctly answered all items on the task. Internal consistency estimates, test-retest estimates, difficulty indices, and item-to-total correlations were not reported since they were poorly estimated due to the lack of variability in scores. Using a mastery criterion of 4 or more

correct, 97% of students achieved mastery on Occasion 1 and 97% on Occasion 2. Raters agreed on 35 of 36 students, all 35 of whom scored above criterion.

4. **Blending Phonemes** - This 5-item task was administered to 35 first grade students on both assessment occasions, 5 only on the first occasion and 9 only on the second occasion. The distribution of total scores was negatively skewed and demonstrated that 81% of all students scored 5 on the task. Internal consistency estimates, test-retest estimates, difficulty indices, and item-to-total correlations were not reported since they were poorly estimated due to the lack of variability in scores. Using a mastery criterion of 4 or more correct, 100% of students achieved mastery on Occasion 1 and 97% on Occasion 2. Raters agreed on 34 of 35 students, all 34 of whom scored above criterion.

5. **Deleting Initial Sounds** - This 5-item task was administered to 34 first grade students on both assessment occasions, 6 only on the first occasion, and 7 only on the second occasion. The distribution of total scores was negatively skewed and demonstrated that 86% of all students scored 5 on the task. Internal consistency was excellent at each occasion ($\alpha=.91$ for Occasion 1 and $\alpha=.86$ for Occasion 2). The test-retest correlation ($r=.88$) demonstrated that scores varied little over the week between occasions and over the two sets of raters. Items in this task were very easy for the students (% correct ranged from .88 -.97). Item-to-total correlations ranged from good to excellent (.39 to .92). Using a mastery criterion of 4 or more correct, 91% of students achieved mastery on both occasions. The dependability index for this task was high, but was only estimable using a technique which assumes normality of scores. Therefore, the index is only a rough estimate ($\phi = .95$). Agreement between raters was excellent as evidenced by kappa=1.0. Raters agreed on all 34 students, 31 of whom scored above criterion. This task correlated most highly with the Initial Blend Substitution Task ($r=.56$) which is not unsurprising since both of these tasks require manipulation of initial phonemes.

Table 28.
Reliability Information for the TPRI End of First Grade Form

	Alpha Time 1	Alpha Time 2	*Test-Retest	Generalizability Coefficient	Test Difficulty Time 1	Test Difficulty Time 2	Kappa for Task
SCREEN							
Word Reading	.88	.89	.93	.95	.09	.16	.69
Blending Phonemes	.74	.75	.51	.58	.16	.20	.87
INVENTORY							
Book and Print Awareness	N/A	-.06	.36	.55	1.0	1.0	N/A
Rhyming	.83	.80	.75	.98	.80	.93	.26
Blending Onset Rimes	.77	.28	-.06	0	.97	.97	-.03
Blending Phonemes	-.11	.57	.14	.62	1.0	.97	0
Deleting Initial Sounds	.91	.86	.88	.95	.91	.91	1.0
Deleting Final Sounds	.75	.84	.53	.81	.74	.77	.56
Initial Consonant Substitution	.94	N/A	N/A	N/A	.98	1.0	0
Final Consonant Substitution	.96	-.03	-.02	N/A	1.0	1.0	N/A
Medial Vowel Substitution	.75	.65	.47	.89	.91	.86	.32
Initial Blend Substitution	.71	.52	.32	.81	.92	.86	.16
Final Blend Substitution	.87	-.06	.42	.95	.97	1.0	0
Comprehension 1	.46	.40	.45	.56	.66	.77	.51
Comprehension 2	.66	.59	.49	.66	.27	.48	.40

Note.

* Test –Retest reliability is a lower-bound estimate, since two different raters were used. One week passed between the two occasions.

** Sample sizes for both occasions fell below 20.

N/A – not estimable due to limited variability on this task.

6. **Deleting Final Sounds** - This 5-item task was administered to 31 first grade students on both assessment occasions, 5 only on the first occasion, and 8 only on the second occasion. The distribution of total scores was negatively skewed and demonstrated that 47% of all students scored 5 on the task. Internal consistency estimates were good to excellent at each occasion ($\alpha=.75$ for Occasion 1 and $\alpha=.84$ for Occasion 2). The test-retest correlation ($r=.53$) demonstrated that scores varied some over the week between occasions and over the two sets of raters. Items in this task ranged from easy to very easy for the students (% correct ranged from .71 -.90). Item-to-total correlations ranged from adequate to excellent (.27 to .77). Using a mastery criterion of 4 or more correct, 74% of students achieved mastery on Occasion 1 and 77% on Occasion 2. The dependability index for this task indicated that differences between students accounted for most of the variability in scores ($\phi = .81$). Agreement between raters was fair as evidenced by kappa= .56. Raters scored 21 students above criterion on both occasions, 5 students below criterion on both occasions, 2 students above criterion only on the first occasion and 3 students above criterion only on the second occasion. This task correlated most highly with the Word Reading Task ($r=.57$) which seems reasonable since this task assesses phonemic awareness skills which have been shown to highly relate to word reading.
7. **Initial Consonant Substitution** - This 5-item task was administered to 44 first grade students on both assessment occasions, 4 only on the first occasion, and 4 only on the second occasion. The distribution of total scores demonstrated that 99% of all students scored a perfect 5 on the task. Internal consistency estimates, test-retest estimates, difficulty indices, and item-to-total correlations were not reported since they were poorly estimated due to the lack of variability in scores. Using a mastery criterion of 4 or more correct, 98% of students achieved mastery on Occasion 1 and 100% on Occasion 2. Raters agreed on 43 of 44 students, with all 43 scoring above criterion.
8. **Final Consonant Substitution** - This 5-item task was administered to 43 first grade students on both assessment occasions, 4 only on the first occasion, and 5 only on the second occasion. The distribution of total scores demonstrated that 96% of all students scored a perfect 5 on the task. Internal consistency estimates, test-retest estimates, difficulty indices, and item-to-total correlations were not reported since they were poorly estimated due to the lack of variability in scores. Using a mastery criterion of 4 or more correct, 98% of students achieved mastery on Occasion 1 and 100% on Occasion 2. Raters scored all students above criterion on both occasions.
9. **Medial Vowel Substitution** - This 5-item task was administered to 43 first grade students on both assessment occasions, 4 only on the first occasion, and 5 only on the second occasion. The distribution of total scores was negatively skewed and demonstrated that 73% of students scored a 5 on the task. Internal consistency estimates were good at each occasion ($\alpha=.75$ for Occasion 1 and $\alpha=.65$ for Occasion 2). The test-retest correlation ($r=.47$) demonstrated that scores varied some over the week between occasions and over the two sets of raters. Items in this task were very easy for the students (% correct ranged from .84 -.98). Item-to-total correlations ranged from poor to excellent (.16 to .77). Using a mastery criterion of 4 or more correct, 91% of students achieved mastery on Occasion 1 and 86% on

Occasion 2. The dependability index for this task indicated that differences between students accounted for most of the variability in scores ($\phi = .89$). Nevertheless, agreement between raters on mastery decisions was poor as evidenced by $\kappa = .32$; however agreement should improve with further training on this task. Raters agreed on 37 of 43 students, with 35 scoring above criterion on both occasions. Of the 6 disagreements, 4 students were scored above criterion only on the first occasion. This task correlated most highly with the Deleting Final Sounds and Deleting Initial Sounds Tasks ($r = .45$ and $r = .44$, respectively).

10. **Initial Blend Substitution** - This 5-item task was administered to 36 first grade students on both assessment occasions, 7 only on the first occasion, and 6 only on the second occasion. The distribution of total scores was negatively skewed and demonstrated that 73% of students scored a 5 on the task. Internal consistency was good at each occasion ($\alpha = .71$ for Occasion 1 and $\alpha = .52$ for Occasion 2). The test-retest correlation ($r = .32$) demonstrated that scores varied considerably over the week between occasions and over the two sets of raters. Items in this task ranged from easy to very easy for the students (% correct ranged from .72 - .97). Item-to-total correlations ranged from adequate to excellent (.35 to .67). Using a mastery criterion of 4 or more correct, 92% of students achieved mastery on Occasion 1 and 86% on Occasion 2. The dependability index for this task indicated that differences between students accounted for much of the variability in scores ($\phi = .81$). Nevertheless, agreement between raters for mastery decision was poor as evidenced by $\kappa = .16$; however agreement should improve with further training on this task. Raters scored 29 students above criterion on both occasions, 1 student below criterion on both occasions, 4 students above criterion only on the first occasion and 2 students above criterion only on the second occasion. This task correlated most highly with the Final Blend Substitution Task ($r = .59$).
11. **Final Blend Substitution** - This 5-item task was administered to 31 first grade students on both assessment occasions, 10 only on the first occasion, and 6 only on the second occasion. The distribution of total scores was negatively skewed and demonstrated that 86% of students scored a 5 on the task. Internal consistency was excellent at the first occasion ($\alpha = .94$). At Occasion 2, all students correctly answered 3 items, so variability existed in only 2 items. Because of this, the internal consistency estimate was very low ($\alpha = -.06$ for Occasion 2). The test-retest correlation ($r = .42$) demonstrated that scores varied some over the week between occasions and over the two sets of raters. Items in this task were very easy for the students (% correct ranged from .94 - 1.0). Item-to-total correlations at Occasion 1 ranged from adequate to excellent (.42 to .87). At Occasion 2, the 3 items with no variability prevented estimation of item-to-total correlations. Using a mastery criterion of 4 or more correct, 97% of students achieved mastery on Occasion 1 and 100% on Occasion 2. The dependability index for this task was only estimable using a technique which assumes normality of scores, therefore the index is only a rough estimate ($\phi = .95$). Because all students except one met mastery at both occasions and agreement between raters takes chance into account, agreement between raters was poor as evidenced by $\kappa = 0$. Raters agreed on 30 of 31 students, all of whom scored above the criterion. The 1 disagreement was a student who scored above criterion only on the second occasion. This task correlated most highly with the Final Consonant Substitution Task ($r = .85$).

12. **Comprehension 1** - This 5-item task was administered to 44 first grade students on both assessment occasions, 4 only on the first occasion, and 4 only on the second occasion. The distribution of total scores was negatively skewed and demonstrated that 72% of all students scored 4-5 on the task. Internal consistency estimates at each occasion were not as high as expected ($\alpha=.46$ for Occasion 1 and $\alpha=.40$ for Occasion 2). The low internal consistency estimates may reflect that some questions were implicit and others explicit. The test-retest correlation ($r=.45$) demonstrated that scores varied some over the week between occasions and over the two sets of raters. Items in this task ranged from difficult to very easy for the students (% correct ranged from .32 - 1.0). Item-to-total correlations fell in the poor and good ranges (.15 to .47). Using a mastery criterion of 4 or more correct, 66% of students achieved mastery on Occasion 1 and 77% on Occasion 2. The moderate dependability index for this task indicated that differences between students accounted for much of the variability in scores ($\phi = .56$). Agreement between raters was fair as evidenced by $\kappa=.51$. Raters scored 27 students above criterion on both occasions, 8 students below criterion on both occasions, 2 students above criterion on the first occasion only, and 7 students above criterion only on the second occasion. This task correlated most highly with the Comprehension 2 task ($r=.56$).
13. **Comprehension 2** - This 5-item task was administered to 44 first grade students on both assessment occasions, 4 only on the first occasion, and 4 only on the second occasion. The distribution of total scores was rectangular-shaped and demonstrated that the most common score was a 2. Internal consistency was good at each occasion ($\alpha=.66$ for Occasion 1 and $\alpha=.59$ for Occasion 2). The test-retest correlation ($r=.49$) demonstrated that scores varied some over the week between occasions and over the two sets of raters. Items in this task ranged from difficult to easy for the students (% correct ranged from .34 - 70). Item-to-total correlations fell in the poor and excellent ranges (.21 to .62). Using a mastery criterion of 4 or more correct, 27% of students achieved mastery on Occasion 1 and 48% on Occasion 2. The moderate dependability index for this task indicated that differences between students accounted for much of the variability in scores ($\phi = .66$). Agreement between raters was fair as evidenced by $\kappa=.41$. Raters scored 10 students above criterion on both occasions, 21 students below criterion on both occasions, 2 students above criterion only on the first occasion, and 11 students above criterion only on the second occasion. This task correlated most highly with the first comprehension task ($r=.56$).

Beginning Grade 2 Analysis

Table 29 summarizes the reliability information for the Beginning Grade 2 screen and Grade 2 inventory. The Word Reading screen task shows uniformly excellent reliabilities (medians .88-.91). For the inventory, the median estimates were in the good range: alpha (.67), test-retest (.65), and generalizability (.76). Weaknesses in some of the phonological awareness and spelling tasks reflect the difficulty levels of the items. Many tasks are too easy and will require revision or dropping. A detailed discussion follows.

Reliability of Grade 2 Screen

1. **Word Reading Task** - This 7-item task was administered to 62 first grade students on both assessment occasions, 9 only on the first occasion, and 7 only on the second occasion. The distribution of total scores was positively skewed and demonstrated that 58% of all students scored 0 on the task. Internal consistency was excellent at each occasion ($\alpha=.92$ for Occasion 1 and $\alpha=.91$ for Occasion 2). The test-retest correlation ($r=.90$) demonstrated that scores varied little over the week between occasions and over the two sets of raters. Most items in this task ranged from very difficult to difficult for the students (% correct ranged from .15 - .31). Item-to-total correlations ranged from good to excellent (.44 to .89). Using a mastery criterion of 5 or more correct, 21% of students achieved mastery on Occasion 1 and 24% on Occasion 2. The high dependability index for this task indicated that differences between students accounted for most of the variability in scores ($\phi = .88$). Agreement between raters was excellent as evidenced by kappa=.82. Raters scored 9 students above criterion on both occasions, 48 students below criterion on both occasions, 2 students above criterion only on the first occasion, and 3 students above criterion only on the second occasion. This task correlated most highly with the Spelling of Long Vowels Task ($r=.72$).

Reliability of Grade 2 Inventory

1. **Initial Consonant Substitution** - This 5-item task was administered to 62 first grade students on both assessment occasions, 9 only on the first occasion, and 7 only on the second occasion. The distribution of total scores was negatively skewed and demonstrated that 93% of all students scored a perfect 5 on the task. Internal consistency ranged from good to excellent ($\alpha=.83$ for Occasion 1 and $\alpha=.55$ for Occasion 2). Because so little variability existed in the scores, the test-retest correlation was low ($r= -.03$). Items in this task were very easy for the students (% correct ranged from .97 - 1.0). Item-to-total correlations ranged from good to excellent (.40 to .81). Using a mastery criterion of 4 or more correct, 98% of students achieved mastery on both occasions. Raters scored 60 students above criterion on both occasions, 1 student above criterion only on the first occasion and 1 student above criterion only on the second occasion.
2. **Final Consonant Substitution** - This 5-item task was administered to 60 first grade students on both assessment occasions, 9 only on the first occasion, and 7 only on the second occasion. The distribution of total scores was negatively skewed and demonstrated that 69% of all students scored a perfect 5 on the task. Internal consistency was good at each occasion ($\alpha=.64$ for Occasion 1 and $\alpha=.78$ for Occasion 2). The test-retest correlation ($r=.40$) demonstrated that scores varied some over the week between occasions and over the two sets of raters. Items in this task were very easy for the students (% correct ranged from .83 - .93).

Table 29
Reliability Information for the TPRI Beginning of Second Grade Form

	Alpha Time 1	Alpha Time 2	*Test-Retest	Generalizability Coefficient	Test Difficulty Time 1	Test Difficulty Time 2	Kappa for Task
SCREEN							
Word Reading	.92	.91	.90	.88	.21	.24	.69
INVENTORY							
Initial Consonant Substitution	.83	.55	-.03	.98	.98	.98	-.02
Final Consonant Substitution	.64	.78	.40	.61	.90	.88	.22
Medial Vowel Substitution	.68	.82	.69	.67	.88	.83	.61
Initial Blend Substitution	.85	.79	.27	.66	.77	.82	.16
Final Blend Substitution	.46	.40	-.10	.02	.97	.93	-.05
Spelling CVC and CVCe	.74	.70	.70	.89	.73	.85	.43
**Spelling Long Vowels	.75	.56	.90	.94	.61	.67	.88
**Spelling Past Tense	.44	.41	.90	.96	.55	.73	.62
**Spelling Endings	.24	.58	.86	.87	.50	.50	1.0
Comprehension 1	.66	.70	.65	.61	.24	.44	.59
Comprehension 2	.67	.68	.59	.76	.39	.56	.53

Note.

* Test –Retest reliability estimate is a lower-bound estimate, since two different raters were used. One week passed between the two occasions.

** Sample sizes for both occasions fell below 20.

Item-to-total correlations ranged from poor to excellent (.17 to .68). Using a mastery criterion of 4 or more correct, 90% of students achieved mastery on Occasion 1 and 88% on Occasion 2. The dependability index for this task indicated that differences between students accounted for much of the variability in scores ($\phi = .61$). Agreement between raters was poor as evidenced by kappa= .22. Raters agreed on 51 of 60 students, with 49 students scoring above criterion on both occasions. The 9 disagreements were almost evenly split with 5 students scoring above criterion only on the first occasion and 4 scoring above criterion only on the second occasion. This task correlated most highly with the Medial Vowel Substitution Task ($r=.62$).

3. **Medial Vowel Substitution** - This 5-item task was administered to 52 first grade students on both assessment occasions, 9 only on the first occasion, and 9 only on the second occasion. The distribution of total scores was slightly negatively skewed and demonstrated that 63% of all students scored a 5 on the task. Internal consistency ranged from good to excellent ($\alpha=.68$ for Occasion 1 and $\alpha=.82$ for Occasion 2). The test-retest correlation ($r=.69$) demonstrated that scores varied some over the week between occasions and over the two sets of raters. Items in this task were very easy for the students (% correct ranged from .81 -.90). Item-to-total correlations ranged from adequate to good (.24 to .59). Using a mastery criterion of 4 or more correct, 88% of students achieved mastery on Occasion 1 and 83% on Occasion 2. The dependability index for this task indicated that differences between students accounted for much of the variability in scores ($\phi = .67$). Agreement between raters was good as evidenced by kappa= .61. Raters scored 42 students above criterion on both occasions, 5 students below criterion on both occasions, 4 students above criterion only on the first occasion, and 1 student above criterion only on the second occasion. This task correlated most highly with the Final Consonant Substitution Task ($r=.62$).
4. **Initial Blend Substitution** - This 5-item task was administered to 44 first grade students on both assessment occasions, 11 only on the first occasion, and 8 only on the second occasion. The distribution of total scores was negatively skewed and demonstrated that 55% of students scored a 5 on the task. Internal consistency was excellent at each occasion ($\alpha=.85$ for Occasion 1 and $\alpha=.79$ for Occasion 2). The test-retest correlation ($r= .27$) demonstrated that scores varied considerably over the week between occasions and over the two sets of raters. Items in this task ranged from easy to very easy for the students (% correct ranged from .66 -.91). Item-to-total correlations ranged from good to excellent (.46 to .74). Using a mastery criterion of 4 or more correct, 77% of students achieved mastery on Occasion 1 and 82% on Occasion 2. The dependability index for this task indicated that differences between students accounted for much of the variability in scores ($\phi = .66$). Nevertheless, agreement between raters for mastery decisions was poor as evidenced by kappa= .16, indicating a need for additional training of raters, or the need for additional items. Raters scored 29 students above criterion on both occasions, 3 students below criterion on both occasions, 5 students above criterion only on the first occasion, and 7 students above criterion only on the second occasion.
5. **Final Blend Substitution** - This 5-item task was administered to 29 first grade students on both assessment occasions, 12 only on the first occasion, and 14 only on the second occasion.

The distribution of total scores was negatively skewed and demonstrated that 89% of students scored a 5 on the task, and all students scored at least 3 on the task. Internal consistency was adequate each occasion ($\alpha=.46$ for Occasion 1 and $\alpha=.40$ for Occasion 2). The test-retest correlation ($r= -.10$) demonstrated that scores varied considerably over the week between occasions and over the two sets of raters. Items in this task were very easy for the students (% correct ranged from .93 - 1.0). At each occasion, all students scored a perfect 5 on one item, so no variability in scores existed for one item at each occasion. Item-to-total correlations for items with variability ranged from poor to excellent (-.05 to .61). Using a mastery criterion of 4 or more correct, 97% of students achieved mastery on Occasion 1 and 93% on Occasion 2. The lack of variability in scores prevented estimation of specific variance components, so the dependability index is not reported. Because all but 3 students met mastery at both occasions and agreement between raters takes chance into account, agreement between raters was poor as evidenced by $\kappa=-.05$. This task correlated most highly with the Spelling CVC and CVCe Task ($r=.35$).

6. **Spelling CVC and CVCe Patterns** - This 5-item task was administered to 26 first grade students on both assessment occasions, 14 only on the first occasion, and 15 only on the second occasion. The distribution of total scores was negatively skewed and demonstrated that 43% of students scored a 5 on the task. Internal consistency was good at each occasion ($\alpha=.74$ for Occasion 1 and $\alpha=.70$ for Occasion 2). The test-retest correlation ($r= .70$) demonstrated that scores varied little over the week between occasions and over the two sets of raters. Items in this task ranged from moderately difficult to very easy for the students (% correct ranged from .46 - 1.0). Item-to-total correlations ranged from poor to excellent (.15 to .77). The dependability index for this task indicated that differences between students accounted for most of the variability in scores ($\phi = .89$). Using a mastery criterion of 4 or more correct, 73% of students achieved mastery on Occasion 1 and 85% on Occasion 2. Agreement between raters was adequate as evidenced by $\kappa= .43$. Raters scored 18 students above criterion on both occasions, 3 students below criterion on both occasions, 1 student above criterion only on the first occasion, and 4 students above criterion only on the second occasion. This task correlated most highly with the Spelling Long Vowels Task ($r=.60$).
7. **Spelling Long Vowels** - This 5-item task was administered to 18 first grade students on both assessment occasions, 11 only on the first occasion, and 11 only on the second occasion. The distribution of total scores was negatively skewed and demonstrated that 41% of students scored a 5 on the task. Internal consistency ranged from adequate to good over the two occasions ($\alpha=.75$ for Occasion 1 and $\alpha=.56$ for Occasion 2). The test-retest correlation ($r= .90$) demonstrated that scores varied little over the week between occasions and over the two sets of raters. Items in this task ranged from moderately difficult to very easy for the students (% correct ranged from .50 - 1.0). Item-to-total correlations ranged from poor to excellent (.10 to .84). The dependability index for this task was only estimable using a technique which assumes normality of scores, therefore the index is only a rough estimate ($\phi = .94$). Using a mastery criterion of 4 or more correct, 61% of students achieved mastery on Occasion 1 and 67% on Occasion 2. Agreement between raters was excellent as evidenced by $\kappa= .88$. Raters scored 11 students above criterion on both occasions, 6 students

below criterion on both occasions, and 1 student above criterion only on the second occasion. This task correlated most highly with the Word Reading Task ($r=.72$).

8. **Spelling Orthographic Patterns, Patterns, Conventions, and Past Tense** - This 5-item task was administered to 11 first grade students on both assessment occasions, 7 only on the first occasion, and 4 only on the second occasion. Because of the small sample size, reliability estimates are not considered stable; however, they are reported as preliminary estimates. The distribution of total scores was negatively skewed and demonstrated that 43% of students scored a 5 on the task. Internal consistency was adequate at each occasion ($\alpha=.44$ for Occasion 1 and $\alpha=.41$ for Occasion 2). The test-retest correlation ($r=.90$) demonstrated that scores varied little over the week between occasions and over the two sets of raters. Items in this task ranged from easy to very easy for the students (% correct ranged from .64 - 1.0). Item-to-total correlations ranged from poor to adequate (-.17 to .39). The dependability index for this task was only estimable using a technique which assumes normality of scores, therefore the index is only a rough estimate. Still, the index showed that differences between students accounted for most of the variability in scores ($\phi = .96$). Using a mastery criterion of 4 or more correct, 55% of students achieved mastery on Occasion 1 and 73% on Occasion 2, possibly reflecting repeated exposure. Agreement between raters was good as evidenced by kappa=.62. Raters scored 6 students above criterion on both occasions, 3 students below criterion on both occasions, and 2 students above criterion only on the second occasion. This task correlated most highly with the Spelling Long Vowels Task ($r=.66$).

9. **Spelling Orthographic Patterns, Patterns, Conventions, and Inflectional Endings** - This 5-item task was administered to 6 first grade students on both assessment occasions, 6 only on the first occasion, and 4 only on the second occasion. Because of the small sample size, reliability estimates are not considered stable; however, they are reported as preliminary estimates. The distribution of total scores was approximately normal and demonstrated that 41% of students scored a 3 on the task. Internal consistency ranged from poor to adequate across the two occasions ($\alpha=.24$ for Occasion 1 and $\alpha=.58$ for Occasion 2). The test-retest correlation ($r=.86$) demonstrated that scores varied little over the week between occasions and over the two sets of raters. Items in this task ranged from very difficult to very easy for the students (% correct ranged from .17 - 1.0). Item-to-total correlations were in the poor to excellent range (-.09 to .68). The dependability index for this task was only estimable using a technique which assumes normality of scores, therefore the index is only a rough estimate. Nevertheless, differences between students accounted for most of the variability in scores ($\phi = .87$). Using a mastery criterion of 4 or more correct, 50% of students achieved mastery on Occasion 1 and 50% on Occasion 2. Agreement between raters was perfect as evidenced by kappa= 1.0. Raters scored 3 students above criterion and 3 students below criterion on both occasions. This task correlated most highly with the Spelling CVC and CVCe Task ($r=.57$).

10. **Comprehension 1** - This 5-item task was administered to 62 first grade students on both assessment occasions, 9 only on the first occasion, and 7 only on the second occasion. The distribution of total scores was approximately normal and demonstrated that 25% of all students scored 2 on the task. Internal consistency estimates at each occasion were good ($\alpha=.66$ for Occasion 1 and $\alpha=.70$ for Occasion 2). The test-retest correlation ($r=.65$)

demonstrated that scores did not vary much over the week between occasions and over the two sets of raters. Items in this task ranged from difficult to very easy for the students (% correct ranged from .24 - .84). Item-to-total correlations fell in the adequate to good range (.31 to .58). Using a mastery criterion of 4 or more correct, 24% of students achieved mastery on Occasion 1 and 44% on Occasion 2, possibly reflecting repeated exposure. The dependability index for this task indicated that differences between students accounted for much of the variability in scores ($\phi = .61$). Agreement between raters was adequate as evidenced by kappa=.59. Raters scored 15 students above criterion on both occasions and 35 students below criterion on both occasions, but disagreed on 12 students, all of whom scored above criterion only on the second occasion.. This task correlated most highly with the Comprehension 2 Task ($r=.62$).

- 11. Comprehension 2** - This 5-item task was administered to 62 first grade students on both assessment occasions, 9 only on the first occasion, and 7 only on the second occasion. The distribution of total scores was negatively skewed and demonstrated that 26% of all students scored 5 on the task. Internal consistency was good at each occasion ($\alpha=.67$ for Occasion 1 and $\alpha=.68$ for Occasion 2). The test-retest correlation ($r=.59$) demonstrated that scores varied some over the week between occasions and over the two sets of raters. Items in this task ranged from difficult to easy for the students (% correct ranged from .37 - .76). Item-to-total correlations ranged from poor to excellent (.22 to .60). Using a mastery criterion of 4 or more correct, 39% of students achieved mastery on Occasion 1 and 56% on Occasion 2, possibly reflecting repeated exposure. The dependability index for this task indicated that differences between students accounted for most of the variability in scores ($\phi = .76$). Agreement between raters was adequate as evidenced by kappa=.53. Raters scored 22 students above criterion on both occasions, 25 students below criterion on both occasions, 2 students above criterion only on the first occasion, and 13 students above criterion only on the second occasion. This task correlated most highly with the first comprehension task ($r=.62$).

Validity Analyses

Content Validity

Decisions to include specific tasks were made based upon past research in reading development. Tasks were designed to assess students on skills that have been empirically shown to relate to and predict reading.

Graphophonemic knowledge

Readers have been shown to develop letter-name identification and letter-sound identification skills early in the reading process. Treiman et al. (in press) reported that students learn letter-names first and their knowledge of letter-names facilitates the learning of letter-sounds. Furthermore, letter-name and letter-sound identification skills are necessary for reading (Vellutino et al., 1996), because they advance the development of both orthographic and phonological processing skills. Before readers are able to associate and decode multiple grapheme-phoneme patterns, they learn to make letter-sound associations at the single grapheme-phoneme level. Graphophonemic knowledge skills were included on each form of the TPRI.

Phonological awareness

Research over the past 30 years provides converging evidence that poor reading ability is caused by poor word recognition skills, which in turn is associated with poor development of phonological awareness skills (Lyon, 1995; Perfetti, 1985; Stanovich, 1991; Torgesen, Wagner, Rashotte, Alexander, & Conway, 1997). Thus, students who are unable to apply the alphabetic principle are typically unable to associate phonemes with the corresponding graphemes. A student unable to identify sounds associated with letters does not have the necessary skills to decode common words and new words. Furthermore, when beginning to read sentences and paragraphs, the poor reader who is unable to associate sounds with letters spends so much energy attempting to decode words that comprehension is impaired.

Researchers refer to phonological processing as the mental processes used when individuals manipulate the sound structure of language. A prerequisite skill for phonological processing is phonological awareness, or the awareness of the sound structure of oral language. Students must first be aware of sounds before they are able to apply those sounds to letters. This is a metacognitive skill reflecting a superordinate awareness of the sound structures of oral language.

Researchers have reported a strong relation between phonological processing and word reading (Calfee, Lindamood, & Lindamood, 1973; Fletcher et al., 1994; Rosner & Simon, 1971; Shaywitz, Fletcher, Holahan, & Shaywitz, 1992; Stanovich & Siegel, 1994). A body of evidence has evolved suggesting that the relationship is not just correlational, but causal (Mann & Liberman, 1984; Share, Jorm, MacLean, & Matthews, 1984; Torgesen et al., 1994). Because the research has demonstrated an important link between phonological processing and reading, phonological processing skills were included in the TPRI.

Orthographic processing

Orthographic processing refers to the representation of sounds with printed or written symbols. While the relation between orthographic and phonological processing has not proven to be strong at lower levels of reading and orthographic processing (Stanovich & Siegel, 1994), the relation has proven to be strong when orthographic and phonological processing skills are more developed (Foorman et al., 1996). Researchers have demonstrated that orthographic processing shares unique variance with word reading above and beyond phonological processing

(Cunningham & Stanovich, 1990). In addition, word spelling has also been shown to predict word reading (Foorman et al., 1991). Therefore, orthographic processing skills were included in the graphophonemic sections of the TPRI.

Comprehension

The ultimate purpose of learning to read is to enable the child to understand what they are reading. Successful reading extends beyond the development of word recognition skills (Pressley, 1998). Children must be able to understand the material that they decode. Word recognition skills are a necessary prerequisite to fully developed comprehension skills, and both must be taught. The basis for the development of reading comprehension skills lies in the child's early development of strategies for listening comprehension, which have their basis in the child's oral language skills (Pressley, 1998). However, reading comprehension is more than the simple combination of word recognition skills and oral language (Hoover & Gough, 1980). A variety of processes including knowledge, experience, memory, and attention are involved (Fielding & Pearson, 1994; Pressley, 1998).

In teaching reading comprehension, it is apparent that separate processes underlie the ability to understand narrative text (stories) versus expository text (textbooks, magazines). The development of comprehension skills involving narrative text revolves around the child's ability to build and activate background knowledge and to understand the structure of stories (Pearson & Fielding, 1994). In contrast, teaching children to understand expository text usually involves attention to text structure, visual representations, and summarization. Consequently, the comprehension stories in the TPRI were developed to provide assessments of both narrative and expository text. This will permit the teacher to set learning objectives involving either set of skills.

Item Development

The item development process followed five steps. First, items were selected from pools of available items known to measure the specific content domains specified in the preceding section. For all items in the pools, the parameters for reading and reading related items were examined from an Item Response Theory perspective. Items were selected from the pools based on their difficulty and discrimination parameters. Second, new items were generated based on the item parameter information for existing items that matched the new items as closely as possible in terms of stimulus characteristics and response demands. These items were generated so that they ranged in difficulty while still discriminating well at the ability level for which the test was designed. Third, reading experts from a variety of backgrounds provided item-level criticism and suggestions. These experts examined the items and provided feedback on several occasions. Fourth, items were aligned with the Texas Essential Knowledge and Skills (TEKS). Finally, items were refined and included in the TPRI.

After the five item development steps, the TPRI final items were field tested using the 274 students who participated in the field study in a large urban district in Texas (see above). Results from this field test will be used to further refine the TPRI.

Criterion-related concurrent validity

Student performance on the TPRI was related to performance on several concurrent measures of reading and reading-related skills. The concurrent assessments were gathered as part of another larger NICHD-supported study begun in 1997 in a district other than the one used for the 1992-1996 study described in the introduction. Correlations between the TPRI and measures of graphophonemic knowledge, phonemic awareness, spelling, rapid serial naming, and reading comprehension provided evidence that the TPRI measured the same skills as the other, well-established tests of reading and reading-related skills.

Concurrent measures

Correlations between student performance on the teacher administered TPRI and student performance on the tasks administered by trained assessment persons as part of the larger NICHD study are presented in the next section. The Letter-Name and Letter-Sound tasks, both with reliabilities that exceeded $\alpha = .90$, included all 26 letters and their associated sounds. The Blending Onset-Rime, Blending Phonemes, Blending Phonemes of Nonwords, Blending Phonemes of Words, Segmenting Phonemes, Sound Categorization, and Phoneme Elision measures were derived from the CTPP (Torgesen et al., in press). All of the tasks from the CTPP battery have reported internal consistency estimates that exceed .70.

On Rapid Automatized Naming of Letters and Objects (RNL and RNO) tests (Denckla & Rudel, 1976), students were required to rapidly name familiar objects, and letters as quickly as possible. This task was designed to assess automaticity of serial naming skills. Test-retest reliability was reported from kindergarten to Grade one as .57. However, over this time span, true individual change probably contributes to variability in scores (Wolf, Bally, & Morris, 1986). From Grade one to two, test-retest correlation was estimated at .77.

As a measure of receptive vocabulary, the Peabody Picture Vocabulary Test-Revised (PPVT-R; Dunn & Dunn, 1981) was administered. The PPVT-R has been shown to be highly predictive of reading achievement (Satz, Taylor, Friel, & Fletcher, 1978). The median split-half reliability estimate for the PPVT exceeds .80.

Decoding skills were measured in the NICHD study with the Woodcock-Johnson Psycho-educational Test Battery-Revised (WJR; Woodcock & Johnson, 1989). The Basic Reading composite, composed of the Word Identification and Word Attack tasks, was obtained from Kindergarten students. Internal consistency estimates for the Word Identification and Word Attack tasks were reported as $\alpha=.96$ and $\alpha=.95$, respectively. Rate and accuracy of oral reading and comprehension of text, were measured with the Gray Oral Reading Test-III (Wiederholt, 1994).

End of kindergarten analysis

The correlations in Table 30 demonstrate criterion-related validity evidence for the TPRI end of kindergarten tasks. The TPRI Letter-Name task correlated most highly with the Letter-Name, Letter-Sound, and Rapid Naming of Letters tasks from the NICHD study. Since all of these tasks include letter identification skills, the high correlations provide evidence that they measure similar skills. In addition, the TPRI Letter-Sound task correlated most highly with the Letter-Sound, Rapid Naming of Letters, and Sound Categorization tasks. Again, all of these tasks directly assess letter-sound skills, thus providing evidence of validity. The third TPRI screening task, Blending Onset-Rime, correlated most highly with the Blending Onset-Rime, Blending Phonemes in Words, Letter-Sounds, and Sound Categorization tasks. These high correlations support the validity of the TPRI Blending Onset-Rime task, since all of these tasks require students to identify and manipulate the sounds of single phonemes. Additionally, all of these tasks except the Blending Phonemes in Words focus on initial phonemes in words.

The TPRI Book and Print Awareness and Rhyming tasks correlated only moderately with the concurrent reading measures. The moderate correlations were expected since these tasks demonstrated the lowest reliability of the TPRI kindergarten tasks. The TPRI Blending Word Parts and Blending Phonemes tasks correlated most highly with the Letter-Sound, Blending Onset-Rime, and Sound Categorization tasks. Since these tasks focus on manipulating sounds, the correlations suggest that they measure similar skills. The TPRI Deleting Initial Sounds task correlated most highly with the phonemic awareness tasks from the CTPP battery. This supports the validity of the Deleting Initial Sounds task, since the Deleting Initial Sounds task was designed to assess phonemic awareness skills. The Deleting Final Sounds task correlated most highly with the Letter-Sounds, Phoneme Elision, and WJR Basic Reading tasks. The high correlation with the Letter-Sounds task was expected, since both the Deleting Final Sounds and Letter-sounds tasks focus on single phonemes. Both the TPRI Deleting Final Sounds and Phoneme Elision tasks require students to delete single phonemes, so the high correlation provided evidence of validity.

The TPRI Letter-name Identification task correlated most highly with the Letter-Name, Letter-Sound, Rapid Naming of Letters, and PPVT-R tasks. While the high correlations with the Letter-Name, Letter-Sound, and Rapid Naming of Letters tasks were expected, the high correlation with the PPVT-R was not surprising since both the identification of letters names and the recognition vocabulary have been shown to predict reading achievement. Both the TPRI Letter-to-Sound Linking tasks focus on single phoneme sounds, so the high correlations with the Letter-Sounds, Rapid Naming of Letters, and Sound Categorization tasks provide evidence of concurrent validity.

Table 30
 Concurrent Validity for the TPRI End of Kindergarten Form

	Graphophonemic Tasks		Phonemic Awareness Criterion Tasks						Serial Naming Tasks		Other Tasks	
	LN	LS	BOR	BPN	BPW	PE	SC	SPH	RNL	RNO	PV	WJR
SCREEN												
Letter Name	.95*	.67*	.33	.29	.35	.19	.48*	.22	.86*	.48*	.51*	.24
Letter Sound	.64*	.86*	.57*	.51*	.59*	.39*	.76*	.42*	.59*	.54*	.59*	.44*
Blending Onset Rime	.40*	.65*	.69*	.58*	.67*	.44*	.70*	.39*	.46*	.48*	.54*	.45*
INVENTORY												
Book and Print Awareness	.27	.51*	.32*	.34	.41*	.32	.47*	.32	.30	.33	.50*	.35
Rhyming	.25	.26	.21	.10	.25	.09	.29	.13	.32	.29	.46*	.19
Blending Word Parts	.41*	.76*	.67*	.56*	.60*	.41*	.69*	.43*	.45*	.41*	.47*	.36
Blending Phonemes	.43*	.75*	.60*	.46*	.49*	.32	.72*	.36	.50*	.53*	.52*	.34
Deleting Initial Sounds	.40*	.64*	.73*	.75*	.77*	.61*	.61*	.65*	.50*	.54*	.70*	.55*
Deleting Final Sounds	.42*	.53*	.48*	.49*	.56*	.58*	.43*	.14	.57*	.55*	.39*	.52*
Letter Name Identification	.88*	.75*	.35	.30	.35	.34	.54*	.23	.85*	.55*	.62*	.26
Letter-to-sound Linking A	.53*	.76*	.33	.25	.36	.27	.66*	.28	.54*	.55*	.47*	.33
Letter-to-sound Linking B	.52*	.81*	.37	.33	.38	.31	.67*	.30	.62*	.46*	.48*	.33
COMPREHENSION 1	.09	.23	.23	.21	.30	.11	.26	.19	.17	.36	.41*	.22
COMPREHENSION 2	.33	.50*	.20	.27	.33	.29	.47*	.27	.41*	.57*	.65*	.32
COMPREHENSION 3	.21	.40*	.26	.29	.37	.29	.39*	.29	.24	.38	.63*	.27

Note. LN = Letter Name, LS = Letter Sounds, BOR= Blending Onset-Rime, BPN = Blending Phonemes of Nonwords, BPW = Blending Phonemes of Words, PE = Phoneme Elision, SC = Sound Categorization, SPH = Segmenting Phonemes, RNL = Rapid Naming of Letters, RNO = Rapid Naming of Objects, PV = Peabody Picture Vocabulary Test-Revised, WJR = Woodcock Johnson

* $p < .01$.

The first TPRI comprehension story involved the reading of a passage from Danny and the Dinosaur, a familiar story for many students. The story demonstrated poor to adequate reliability, which may be due to the fact that many of the students had heard the story prior to hearing it with the TPRI. Since reliability is necessary for validity, the low correlations between this first TPRI comprehension task and the criterion measures were expected. The second and third comprehension tasks used passages from less popular books, and the reliability for both of these comprehension stories was adequate. Evidence of validity for the second and third comprehension stories was found, since these two comprehension stories correlated most highly with the PPVT-R. On the three explicit items from both comprehension stories, students were required to recall and recognize details from the story. The ability to recognize details and to recall words related to the details remained similar on both the TPRI comprehension stories and the PPVT-R.

Beginning Grade 1 Analysis

Table 31 presents correlations that demonstrate criterion-related validity evidence for the TPRI beginning of the year first grade screening tasks. The TPRI Letter-name task correlated most highly with the Letter-sounds, and Rapid Naming of Letters tasks. The 10 letters that were included on the TPRI Letter-name task tend to be more difficult; therefore, the moderate correlation with the Letter-name task from the NICHD study was not unexpected. The TPRI Letter-sound task correlated most highly with the Letter-sound, Blending Phonemes of Nonwords, and Blending Phonemes of Words tasks. All of these tasks focus on single letter-sounds, so the high correlations demonstrate evidence of concurrent validity. The TPRI Word Reading task correlated most highly with the WJR Basic Reading Score providing evidence of validity for the Word Reading task. The TPRI Blending Phonemes task correlated most highly with the Blending Phonemes into Words, Blending Onset-Rime, and Blending Phonemes into Nonwords. Since all of these tasks require students to blend phonemes, the high correlations provide evidence of construct validity.

Correlations between TPRI beginning of first grade inventory tasks and tasks from the larger study provide validity evidence for the inventory tasks. Moderate to low correlations between the TPRI Book and Print Awareness and the validation measures were not unexpected, since the TPRI Book and Print Awareness task demonstrated only moderate reliability. The TPRI Rhyming task demonstrated high reliability, yet did not correlate highly with the validation measures. The moderate to low correlations indicate a need for revision of this task. The TPRI Blending Onset-Rime task correlated most highly with the Blending Phonemes into Words, Sound Categorization, and Blending Onset-Rime tasks. All of these phonemic awareness tasks require students to manipulate phonemes, so the high correlations demonstrate validity evidence. The TPRI Blending Phonemes task correlated above $r=.50$ with all of the blending tasks from the validation measures which supports the validity for the TPRI Blending Phonemes task. The high correlations between both the TPRI Deleting Initial Sounds and Deleting Final Sounds tasks with the Phoneme Elision task from the CTPP battery provided validity evidence for both of the TPRI Deleting Sounds tasks.

The TPRI Initial Consonant Substitution task correlated most highly with the Letter-Sounds and the Rapid Naming of Letters tasks. Since all of these tasks require students to identify single

letters and letter-sounds, the pattern of correlations demonstrates evidence of construct validity. The TPRI Final Consonant Substitution task demonstrated only moderate reliability, so the moderate correlations with the validation measures were expected. Even though the correlations were moderate, the pattern of correlations seemed logical. For example, the TPRI Final Consonant Substitution correlated most highly with tasks that focused on single phonemes, and correlated lowest with the Rapid Naming of Objects task, one task that does not focus on phonemes. Evidence for validity for the TPRI Medial Vowel Substitution task was found in the high correlations with the Blending Phonemes in Words and Blending Phonemes in Nonwords. All of these tasks assess students' ability to manipulate single phonemes. Both the TPRI Initial Blend Substitution and Final Blend Substitution tasks demonstrated only moderate reliability, so their correlations with the validation measures were only moderate as well. Even so, the pattern of correlations included higher correlations with graphophonemic and phonemic awareness tasks and lower correlations with tasks that do not assess graphophonemic or phonemic awareness skills.

Both of the TPRI comprehension tasks correlated most highly with the PPVT-R. This result was expected and provided support for the validity of the comprehension stories. In addition, the lower correlations between the TPRI comprehension stories and both the phonemic awareness and the naming tasks were expected, since the comprehension stories were not designed to measure the same skills as the phonemic awareness task and the naming tasks.

End of Grade 1 analysis

The correlations in Table 32 demonstrate construct validity evidence for the TPRI end of the year Grade 1 screen. The TPRI Word Reading screening task correlated most highly with the WJR Word Identification, Word Reading, and KTEA Spelling measures. These high correlations support the validity of the TPRI Word Reading task. However, the correlation between the TPRI Word Reading task and the WJR Word Attack task was only $r = .33$, which was statistically significant, but somewhat lower than expected. The lower correlation may reflect the use of high frequency words. The TPRI Blending Phonemes screening task correlated most highly with the Blending Phonemes into Words, Blending Onset-Rime, Phoneme Elision, and WJR Letter Word Identification tasks. All of these tasks assess a student's ability to manipulate components of words involving sounds, so the high correlations provide evidence of construct validity.

Intercorrelations between the TPRI inventory tasks and the validation measures also provide support for the validity of the TPRI inventory tasks. The moderate to low correlations between the TPRI Book and Print Awareness and the validation measures were not unexpected, since the TPRI Book and Print Awareness task only demonstrated moderate reliability. The correlation between the TPRI Rhyming task and the Blending Onset-Rime task was high, indicating that both tasks assess similar skills. The TPRI Blending Onset-Rime task correlated most highly with the Blending Onset-Rime task from the NICHD study supporting the validity for the TPRI task.

Table 31
 Concurrent Validity for the TPRI Beginning Grade 1 Analysis

	Graphophonemic Tasks		Phonemic Awareness Criterion Tasks						Serial Naming Tasks		Other Tests	
	LN	LS	BOR	BPN	BPW	PE	SPH	SC	RNL	RNO	PV	WJ
SCREEN												
Letter Name	.81*	.71*	.32	.35	.31	.22	.28	.35	.67*	.39*	.37*	.24
Letter Sound	.48*	.71*	.61*	.61*	.61*	.45*	.56*	.56*	.45*	.31	.57*	.52*
Word Reading	.14	.28	.33	.25	.57*	.61*	.47*	.32	.29	.20	.37*	.83*
Blending Phonemes	.38*	.51*	.70*	.64	.74*	.48*	.48*	.57*	.37	.20	.44*	.56*
INVENTORY												
Book and Print Awareness	.30	.39*	.24	.27	.33	.32	.35	.22	.23	.46*	.46*	.28
Rhyming	.38*	.34	.32	.33	.17	.12	.23	.31	.19	.10	.35	.09
Blending Onset-Rime	.33	.52*	.60*	.61*	.65*	.44*	.44*	.55*	.25	.16	.43*	.44*
Blending Phonemes	.39*	.62*	.61*	.61*	.54*	.40*	.48*	.40*	.31	.155	.44*	.35
Deleting Initial Sounds	.11	.32	.39*	.44*	.52*	.58*	.41*	.29	.25	.20	.39*	.43*
Deleting Final Sounds	.12	.37	.36	.56*	.45*	.53*	.38*	.28	.22	.31	.48*	.44*
Initial Consonant Substitution	.56*	.54*	.45*	.46*	.41*	.21	.31	.49*	.48*	.22	.30	.30
Final Consonant Substitution	.46*	.41*	.41*	.45*	.33	.24	.33	.27	.22	-.04	.23	.23
Medial Vowel Substitution	.23	.22	.48*	.57*	.53*	.39*	.44*	.33	.15	.16	.35	.30
Initial Blend Substitution	.28	.26	.25	.47*	.31	.47*	.28	.33	.26	.18	.56*	.36
Final Blend Substitution	.28	.27	.36	.57*	.43*	.47*	.49*	.21	.31	.31	.41*	.336
COMPREHENSION 1	.15	.26	-.08	.13	.11	.14	.09	.01	.20	.31	.38*	.06
COMPREHENSION 2	.17	.31	.22	.20	.32	.233	.29	.26	.22	.20	.47*	.24

Note. LN = Letter Name, LS = Letter Sounds, BOR= Blending Onset-Rime, BPN = Blending Phonemes of Nonwords, BPW = Blending Phonemes of Words, PE = Phoneme Elision, SPH = Segmenting Phonemes, SC = Sound Categorization, RNL = Rapid Naming of Letters, RNO = Rapid Naming of Objects, PV = Peabody Picture Vocabulary Test-Revised, WJ = Woodcock Johnson

* $p < .01$

Due to the lack of variability in items on the Blending Phonemes task, the reliability estimates for this task were poor. Therefore, the correlations between the TPRI Blending Phonemes task and the validation measures were low to moderate.

Both the TPRI Deleting Initial Sounds and Deleting Final Sounds tasks correlated most highly with the Blending Phonemes in Words, Phoneme Elision, and Word Reading tasks. Since all of the tasks are phonemic awareness measures except Word Reading, and since phonemic awareness tasks have been shown to be the best predictors of word reading, these high correlations provide good evidence of concurrent validity.

The TPRI Initial Consonant Substitution, Final Consonant Substitution, and Medial Vowel Substitution tasks correlated most highly with the Blending Onset-Rime and Rapid Naming of Letters tasks. Since the three TPRI tasks focus on letter-sounds, the high correlation with the Blending Onset-Rime task seems logical since the Blending Onset-Rime task also focuses on letter-sounds. The high correlation with the Rapid Naming of Letters also seems reasonable, because both of the TPRI tasks and the Rapid Naming of Letters assess students' ability to identify letters.

Both the TPRI Initial Blend Substitution and Final Blend Substitution tasks correlated most highly with the Blending Onset Rime, Blending Phonemes, and WJR Letter Word Identification tasks. Because all of these tasks require students to manipulate sounds at the phoneme level, the high correlations support the validity of the TPRI tasks.

The first TPRI comprehension story demonstrated only moderate reliability, so the validity evidence was not strong. The moderate reliability and poor validity information suggest a need for further revision of the items for the first TPRI comprehension story. The second comprehension story demonstrated adequate reliability and correlated most highly with the WJR Letter-Word subtest. Furthermore, the second TPRI comprehension story correlated at almost this same level with all of the reading comprehension scores. The pattern of correlations for the second comprehension story with the validation measures provides support for the validity of this measure.

Beginning Grade 2 analysis

Table 33 presents correlations that demonstrate criterion-related validity evidence for the TPRI beginning of the year second grade tasks. The TPRI Word Reading task, the only screening task, correlated highly with almost all of the validation measures except the serial naming tasks. Since the serial naming tasks tap rate factors, and word reading is not timed, a high correlation between serial naming and word reading is not expected. The moderate to high correlations with measures of reading comprehension attests to the strong relationship of word recognition skills and reading comprehension.

Table 32
 Concurrent Validity for the TPRI End of First Grade Form

	Phonemic Awareness Tasks				Word Reading		Naming Tasks		Other Tasks			Comprehension Tasks			
	BOR	BPN	BPW	PE	WR	WJWA	RNL	RNO	PV	WJLW	KTEA	WJPC	G-RT	G-ACC	G-C
SCREEN															
Word Reading	.26	.51*	.46*	.58*	.87*	.66*	.33	.26	.31	.75*	.76*	.71*	.74*	.74*	.51*
Blending Phonemes	.43*	.68*	.63*	.55*	.60*	.50*	.29	.26	.25	.53*	.53*	.47*	.39*	.44*	.19
INVENTORY															
Book and Print Awareness	.52*	.48*	.50*	.42*	.49*	.34	.36	.09	.12	.67*	.40*	.37	.30*	.34*	.20
Rhyming	.67*	.41*	.53*	.39*	.40*	.40*	.36	.46*	.21	.33*	.38	.46*	.37*	.35	.33
Blending Onset-Rime	.84*	.37	.54*	.42*	.30	.30	.49*	.22	-.07	.36	.24	.32	.20	.21	.20
Blending Phonemes	.30	.40*	.36	.58*	.48*	.35	.39	.58*	.09	.50*	.38*	.45*	.38*	.38	.18
Deleting Initial Sounds	.53*	.42*	.66*	.59*	.55*	.44*	.53*	.39	.04	.46*	.40*	.60*	.35*	.39	.38*
Deleting Final Sounds	.31	.34	.45*	.48*	.51*	.45*	.15	.26	.31	.46*	.45*	.53*	.52*	.44*	.30
Initial Consonant Substitution	.70*	.28	.41*	.34	.20	.23	.30	.09	-.01	.31	.21	.31	.12	.12	.14
Final Consonant Substitution	.72*	.30	.43*	.31	.23	.27	.39*	.04	-.01	.32	.21	.33	.15	.15	.19
Medial Vowel Substitution	.44*	.39*	.37	.42*	.43*	.46*	.29	-.03	.38*	.43*	.36	.41*	.30	.32	.41*
Initial Blend Substitution	.48*	.38	.41*	.56*	.44*	.36	.48*	.24	.16	.48*	.34	.43*	.32	.32	.27
Final Blend Substitution	.72*	.47*	.52*	.37	.37	.45*	.24	.56*	.08	.33	.29	.39	.26	.28	.26
COMPREHENSION 1	.26	.01	-.02	.02	.13	-.07	.21	.05	.32	.16	.10	.02	.20	.23	.13
COMPREHENSION 2	.17	.29	.10	.17	.51*	.26	.27	.09	.38	.54*	.41*	.48*	.52*	.51*	.41*

Note. BOR= Blending Onset-Rime, BPN = Blending Phonemes of Nonwords, BPW = Blending Phonemes of Words, PE = Phoneme Elision, WR = Word Reading, WJWA = Woodcock Johnson Word Attack Task, RNL = Rapid Naming of Letters, RNO = Rapid Naming of Objects, PV = Peabody Picture Vocabulary Test-Revised, WJLW = Woodcock Johnson Letter Word Identification Task, KTEA = Kaufman Test of Educational Achievement Spelling Task, WJPC – Woodcock-Johnson Passage Comprehension, G-RT – Gray Oral Reading Test Rate, G-ACC - Gray Oral Reading Test Accuracy, G-C - Gray Oral Reading Test Comprehension

* p < .01

Correlations between the TPRI inventory task and criterion measures provide concurrent validity evidence for the TPRI inventory task. The TPRI Initial Consonant Substitution and Final Consonant Substitution tasks proved to be very easy for the study participants and little variability existed in the tasks, so the reliability of these two TPRI tasks was low.

Validity evidence for the TPRI Initial Blend Substitution and Medial Vowel Substitution tasks was evidenced by the high correlations between the TPRI Initial Blend Substitution and Medial Vowel Substitution tasks and the phonemic awareness and word reading criterion tasks. The TPRI Initial Blend Substitution and Medial Vowel Substitution tasks were designed to assess phonemic awareness, so the high correlations support the validity of the TPRI tasks. The TPRI Final Blend Substitution task items were very easy for students and most students scored a perfect score on the task. The ease of the items resulted in low variability of task scores and poor reliability. The low correlations between the TPRI Final Blend Substitution task and the criterion measures reflected the poor reliability of the task.

The four TPRI spelling tasks correlated most highly with the KTEA Spelling task and the Word Reading task. These high correlations indicate that the TPRI spelling tasks measure similar skills as the KTEA spelling task. Since spelling and word reading have been shown to correlate as students learn to read (Foorman et al., 1991), high correlations between spelling measures and word reading are expected.

The first TPRI reading comprehension story correlated moderately to highly with all of the validation measures, including the reading comprehension measures, with the exception of the serial naming tasks. The low correlation between the first TPRI comprehension passage and the serial naming task is not surprising because this type of naming skill is not strongly related to reading ability in older children without reading problems.

The second TPRI comprehension story did not correlate as highly with the validation measures as the first TPRI comprehension story. While the correlations were not as high, they were significant, particularly with other reading comprehension measures. The pattern of correlations for the second TPRI comprehension story were similar to the pattern of correlations between the first comprehension story and the validation tasks. This similar pattern of correlations provides some evidence of construct validity for the two comprehension stories.

Table 33
 Concurrent Validity for the TPRI Beginning of Second Grade Form

	Phonemic Awareness Tasks				Word Reading		Naming Tasks		Other Tasks			Comprehension Tasks			
	BOR	BPN	BPW	PE	WR	WJWA	RNL	RNO	PV	WJLW	KTEA	WJPC	G-RT	G-ACC	G-C
SCREEN															
Word Reading	.28	.48*	.44*	.50*	.83*	.71*	.49*	.27	.46*	.77*	.84	.63	.75*	.76*	.39*
INVENTORY															
Initial Consonant Substitution	.43*	.33	.43*	.50*	.83*	.71*	.49*	.27	.15	.77*	.30	.38*	.15	.13	.22
Final Consonant Substitution	.39*	.43*	.55*	.55*	.42*	.41*	.46*	.24	.26	.54*	.35	.52*	.26	.21	.44*
Initial Blend Substitution	.50*	.54*	.60*	.47*	.51*	.45*	-.27	.06	.17	.53*	.46*	.58	.40*	.36*	.50
Medial Vowel Substitution	.37	.39	.38	.37	.39	.29	.25	.16	.17	.40	.30	.36*	.23	.24	.33
Final Blend Substitution	.07	.16	.17	.23	.19	.25	-.03	.55*	.10	.17	.16	.22	-.06	-.05	.07
Spelling CVC and CVCe	.0	.14	.20	.29	.70*	.57*	.11	.20	.31	.60*	.65*	.61*	.43*	.48*	.33*
Spelling Long Vowels	-.25	.31	.20	.45*	.84*	.53*	.34	.39	.40*	.73*	.69*	.61*	.62*	.52*	.58*
Spelling Patterns	-.18	-.09	-.26	.09	.77*	.31	-.31	-.06	.42	.61	.53*	.42	.69*	.59*	.11
Spelling Endings	-.09	.23	.04	.29	.28	-.06	.49	-.26	.23	.35	.18	-.27	.51*	.51*	-.31
COMPREHENSION 1	.39*	.48*	.54*	.43*	.61*	.53*	.26	.13	.53*	.57	.53*	.61	.51*	.49*	.56*
COMPREHENSION 2	-.06	.22	.21	.14	.20	.17	.11	-.13	.63*	.17	.08	.26	.27	.23	.36

Note. BOR= Blending Onset-Rime, BPN = Blending Phonemes of Nonwords, BPW = Blending Phonemes of Words, PE = Phoneme Elision, WR = Word Reading, WJWA = Woodcock Johnson Word Attack, RNL = Rapid Naming of Letters, RNO = Rapid Naming of Objects, PV = Peabody Picture Vocabulary Test-Revised, Task, WJLW = Woodcock Johnson Letter Word Identification Task, KTEA = Kaufman Test of Educational Achievement Spelling Task, NWR = Nonword Repetition, WJPC – Woodcock-Johnson Passage Comprehension, G-C - Gray Oral Reading Test Comprehension, G-RT – Gray Oral Reading Test Rate, G-ACC - Gray Oral Reading Test Accuracy

p < .01

Field Study Questionnaire: Teacher Opinions

Thirty teachers completed the TPRI Field study questionnaire on May 11, 1998. Teachers provided their opinions about both training issues and test administration issues.

The CARS staff trained teachers on two occasions. On the first occasion, CARS staff members provided a full-day training for all K-2 teachers in four elementary schools in HISD. The training focused on the purpose and administration of the TPRI. During the second training occasion, the CARS staff provided the 30 teachers who participated in the field study with a half-day of further training focused on administration issues.

The field study questionnaire data is tabulated by question and response in Table 35. Teacher responses are generally very positive for administration with only the Spelling (14%) component rated as "difficult" by more than 3% of teachers. Less than 10% of teachers found the different components of the TPRI to be "average" or only "a little useful." Hence, the majority of teachers were very confident that they administered both the screening portion and the inventory portion correctly. They responded that both the information content and the presentation of the information at both training sessions was very good. Most teachers did not think that training presentations needed to include more multimedia or that they needed more training days. Some of the teachers expressed the need for smaller training groups, fewer speakers during the training, and more time for practice with a trainer. Overall, teachers responded positively to questions about training issues.

97% of teachers felt the TPRI would be helpful (40%) or very helpful (57%) for identifying strengths and weaknesses of the student, while 87% were "likely" (17%) or very likely (70%) to recommend the TPRI to another teacher. 87% were "likely" or "very likely" to recommend the TPRI to an administrator, and 79% were "likely" or "very likely" to recommend the TPRI to the district. In addition, 83% found the directions "easy" or "very easy." Responses about training were also very positive, but many teachers expressed a desire for more training. Their concerns seem to reflect issues involving planning of instruction, where ratings were still positive, but shifted more to the "average" (37%) and "easy" (33%) ratings. 20% of teachers indicated that they felt it would be "a little easy" (7%) or "not easy" (13%) to plan instructional objectives.

Only 40% of teachers indicated familiarity with another reading assessment. Of these teachers, 55-64% felt the TPRI was better than the other assessment, while 33-46% felt the TPRI was comparable. No teachers indicated that the TPRI was worse than another reading assessment.

When asked for suggested changes, 10% felt the screening should be lengthened, while 24% felt it should be shortened. 36% felt the items should be easier, while 8% felt they should be made more difficult. Even in these areas, most teachers (76-92%) felt no changes were needed. For the inventory, 30% felt the content should be shortened, with most other responses were similar to the responses for the screening. The majority of the teachers' comments were directed at training issues (see below), while organizational issues were uniformly positively rated, with at least 70% of teachers rating these parts as "clear" or "very clear." Responses to each question are tabulated in Table 34.

Table 34.
Teacher responses to the TPRI questionnaire.

A. Rate how difficult it was for you to administer each of the following parts of the TPRI.

	N	Very Easy	Easy	Average	Difficult	Very Difficult	N/A
1. Screening Portion	30	47%	33%	13%	3%	3%	0%
2. Book and Print Awareness	30	40%	43%	13%	0%	0%	3%
3. Phonemic Awareness	30	33%	43%	23%	0%	0%	0%
4. Graphophonemic Knowledge	30	37%	37%	27%	0%	0%	0%
5. Spelling	28	25%	32%	11%	14%	0%	18%
6. Comprehension	30	23%	40%	33%	3%	0%	0%

B. Rate how useful you found each of the following parts of the TPRI.

	N	Very Useful	Useful	Average	A Little Useful	Not Useful	N/A
1. Screening Portion	30	47%	33%	3%	7%	10%	0%
2. Book and Print Awareness	30	40%	50%	0%	0%	7%	3%
3. Phonemic Awareness	30	50%	47%	0%	3%	0%	0%
4. Graphophonemic Knowledge	30	47%	53%	0%	0%	0%	0%
5. Spelling	27	37%	41%	4%	0%	0%	19%
6. Comprehension	30	53%	47%	0%	0%	0%	0%

C. Please circle the best response to each question.

1. How helpful do you think the TPRI will be for identifying strengths and weaknesses of students you have not taught?	N	Not At All Helpful	A Little Helpful	Average	Helpful	Very Helpful	N/A
	30	0%	0%	3%	40%	57%	0%
2. Would you recommend the TPRI to another	N	Not At All Likely	A Little Likely	Average	Likely	Very Likely	N/A

teacher?	30	0%	7%	7%	17%	70%	0%
3. Would you recommend the TPRI to an administrator?	N	Not At All Likely	A Little Likely	Average	Likely	Very Likely	N/A
	30	0%	7%	7%	30%	57%	0%
4. Would you recommend the TPRI for district-wide implementation in HISD elementary schools?	N	Not At All Likely	A Little Likely	Average	Likely	Very Likely	N/A
	29	3%	7%	7%	31%	48%	3%
5. How confident are you that you administered the screening portion the way it was intended to be administered?	N	Not At All Confident	A Little Confident	Average	Confident	Very Confident	N/A
	30	3%	0%	10%	33%	53%	0%
6. How confident are you that you administered the inventory portion the way it was intended to be administered?	N	Not At All Confident	A Little Confident	Average	Confident	Very Confident	N/A
	30	0%	0%	13%	40%	57%	0%
7. Describe the <u>information content</u> at the April 21st training.	N	Very Poor	Poor	Average	Good	Very Good	N/A
	29	0%	0%	10%	38%	52%	0%
8. Describe the <u>presentation of information</u> at the April 21st training.	N	Very Poor	Poor	Average	Good	Very Good	N/A
	28	0%	0%	14%	39%	46%	0%
9. How easy were the directions of the TPRI?	N	Very Easy	Easy	Average	A Little Easy	Not Easy	N/A
	30	20%	63%	13%	3%	0%	0%
10. How easy do you think gathering the materials for the TPRI will be?	N	Very Easy	Easy	Average	A Little Easy	Not Easy	N/A
	29	14%	31%	41%	10%	3%	0%

1. Content should be lengthened	29	0%	10%	14%	55%	21%
2. Content should be shortened	29	10%	14%	17%	52%	7%
3. Make items easier	28	7%	29%	4%	54%	7%
4. Make items harder	28	4%	4%	4%	71%	18%
5. Make TPRI easier to score	29	7%	21%	17%	41%	14%
6. Eliminate screening portion	29	14%	7%	10%	59%	10%

F. How do you feel about changing the following aspects of the **inventory portion**?

Suggested Change for Inventory Portion	N	Strongly Agree	Agree	Neither Agree or Disagree	Disagree	Strongly Disagree
1. Content should be lengthened	30	0%	7%	20%	57%	17%
2. Content should be shortened	30	7%	23%	23%	40%	7%
3. Make items easier	29	0%	21%	14%	55%	10%
4. Make items harder	30	3%	7%	13%	67%	10%
5. Make TPRI easier to score	30	0%	20%	37%	37%	7%

G. How do you feel about changing the TPRI training?

Suggested Changes for Training	N	Strongly Agree	Agree	Neither Agree or Disagree	Disagree	Strongly Disagree
1. More multimedia in presentation	29	3%	21%	45%	28%	3%
2. Cover less information	30	0%	10%	37%	43%	10%
3. Smaller groups	28	18%	29%	21%	32%	0%
4. Fewer speakers	29	7%	28%	31%	35%	0%
5. More time to practice with a trainer	30	3%	38%	35%	24%	0%
6. More training days with a shorter amount of time	29	3%	23%	23%	40%	10%
7. Train at a different time of day	29	10%	17%	24%	38%	10%

H. How do you feel about each section of the TPRI?

TPRI Sections	N	Very Confusing	Confusing	Average	Clear	Very Clear	N/A
1. Flow chart	29	0%	0%	24%	48%	17%	10%
2. Directions	30	0%	0%	10%	63%	27%	0%
3. Organization	30	0%	0%	13%	57%	30%	0%
4. Scoring	30	0%	0%	10%	53%	37%	0%
5. Summary Sheets	30	0%	0%	7%	60%	33%	0%
6. Phonemic Awareness Tasks	30	0%	0%	10%	57%	33%	0%
7. Graphophonemic Knowledge Tasks	30	0%	0%	10%	53%	37%	0%
8. Word Reading	30	0%	10%	20%	40%	20%	10%
9. Comprehension	30	0%	7%	10%	53%	30%	0%

Recommended Revisions for Future Development of the TPRI

Based on the reliability and validity data gathered in the TPRI pilot study, several revisions are suggested, most of which involve rewriting of items. Suggested changes target less than adequate reliability for the Book and Print Awareness tasks, a few tasks with an excessive number of easy items, and certain comprehension tasks with passages from familiar books. All of these revisions will be piloted over the next year.

An increase in the reliability of the Book and Print Awareness task might be possible with more explicit directions for administration and more specific descriptions of correct and incorrect responses from students. However, it is also possible that this task is more appropriate for younger children (e.g., four- and five-year-olds). Use of the TPRI does not begin until the middle of kindergarten. The training manual will be revised to address these developmental issues. It may be best to use Book and Print Awareness as a warm-up exercise.

Several tasks were comprised of very easy items, resulting in little variability in scores across these tasks. The lack of variability led to poor reliability estimates for both Phonemic Awareness and Graphophonemic Knowledge tasks, particularly in older children (Grade 2). Writing new items may increase the difficulty levels of these tasks, but they may not be necessary and could potentially be dropped. More specifically, by the end of first grade, items for several phonemic awareness tasks were too easy, including Blending Onset-Rimes, Blending Phonemes, and Deleting Initial Sounds. With respect to the Graphophonemic Knowledge tasks for Grade 1, one

task -- Final Blend Substitution -- was too easy items at both the beginning and end of the year. Initial and Final Consonant Substitution tasks were also too easy by the end of Grade 1. At the beginning of Grade 2, Initial Consonant Substitution and Final Blend Substitution were too easy. Revisions will be made to address these concerns about difficulty levels. New items and a larger set of response options should improve the psychometric properties of the tasks that are presently too easy for the targeted grade levels.

Comprehension passages that were less familiar to students demonstrated the strongest reliability and validity data. For example, the passage from Danny and the Dinosaur, a familiar children's book, demonstrated inadequate reliability. Hence, additional comprehension passages from less familiar books will be added and piloted over the next year.

There is also a need to develop more items for the entire TPRI to reduce effects of familiarity and teaching to the test. For example, the words in the Word Reading screen should be changed periodically. The inclusion of more comprehension passages would give teachers greater flexibility in choosing passages, particularly if the book is familiar to students in the teacher's class. Finally, further examinations of reliability and validity should be completed on larger and more diverse samples. Estimates of item level bias are particularly important, but require larger samples than in the TPRI field study. These studies are in progress.

Teacher and administrator responses to the TPRI in the field study consistently indicated concerns about the need for more training, especially in setting instructional objectives. The need for more training was the only area of concern consistently expressed by teachers. Training should involve not only the administration and scoring of the TPRI, but also intervention strategies. To address training needs, CARS has developed an intervention strategies manual. Information regarding setting learning objectives, along with materials on training, can be obtained from TEA. Over the next year, new training materials will be developed and additional materials concerning instruction around the TPRI results will be forthcoming.

References

- Adams, M.J. (1990). Beginning to read. Cambridge, MA: MIT Press.
- Agronin, M.E., Holahan, J.M., Shaywitz, & Shaywitz, SE (1992). The Multi-Grade Inventory for Teachers. In S.E. Shaywitz & B.A. Shaywitz (Ed.), Attention deficit disorder comes of age (pp. 29-67). Austin, TX: Pro-Ed.
- Beery, K. (1989). Developmental test of visual-motor integration. Chicago: Follett.
- Bowers, P., Golden, J., Kennedy, A., & Young, A. (1994). Limits upon orthographic knowledge due to processes indexed by naming speed. In V.W. Berninger (Ed.). The varieties of orthographic knowledge (Vol. 1, pp. 173-218). Dordrecht, The Netherlands: Kluwer Academic Publishers.
- Bradley, L., & Bryant, P.E. (1983). Categorizing sounds and learning to read – a causal connection. Nature, 301, 419-421.
- Calfee, R. C., Lindamood, P., & Lindamood, C. (1973). Acoustic-phonetic skills and reading - kindergarten through twelfth grade. Journal of Educational Psychology, 64, 293-298.
- Carroll, J., Davies, P., & Richman, B. (1971). The American Heritage word frequency book. Boston, MA: Houghton Mifflin.
- Cunningham, A. E., & Stanovich, K. E. (1990). Assessing print exposure and orthographic processing skill in children: A quick measure of reading exposure. Journal of Educational Psychology, 92, 733-740.
- Cunningham, A.E., & Stanovich, K.E. (1991). Tracking the unique effects of print exposure in children: Associations with vocabulary, general knowledge, and spelling. Journal of Educational Psychology, 83, 264-274.
- Denckla, M. B., & Rudel, R. E. (1976). Naming of objects by dyslexic and other learning-disabled children. Brain and Language, 3, 1-15.
- Dunn, L. M., & Dunn, L. M. (1981). Peabody Picture Vocabulary Test-Revised. Circle Pines, MN. American Guidance Service.
- Dunst, K.J., & Liest, H.E. (1987). Measuring the adequacy of resources in households with young children. Childcare, Health, and Human Development, 13, 111-125.
- Fielding, L.G., & Pearson, P.D. (February, 1994). Reading comprehension: What works. Educational Leadership, 62-68.
- Fletcher, J.M., Shaywitz, S.E., Shankweiler, D.P., Katz, L., Liberman, I.Y., Stuebing, K.K., Francis, D.J., Fowler, A.E., & Shaywitz, B.A. (1994). Cognitive profiles of reading

- disability: Comparisons of discrepancy and low achievement definitions. Journal of Educational Psychology, 86, 6-23.
- Foorman, B. R., Francis, D. J., Novy, D. M., & Liberman, D. (1991). How letter-sound instruction mediates progress in first-grade reading and spelling. Journal of Educational Psychology, 83, 456-469.
- Foorman, B.R., Francis, D.J., Fletcher, J.M., & Lynn (1996). Relation of phonological and orthographic processing to early reading: Comparing two approaches to regression-based, reading-level-match design. Journal of Educational Psychology, 88, 639-652.
- Foorman, B.R., Francis, D.J., Fletcher, J.M., Schatschneider, C., Mehta, P. The role of instruction in learning to read: Preventing reading failure in at-risk-children. Journal of Educational Psychology, 1998, 90, 37-55.
- Francis, D.J., Shaywitz, S.E., Stuebing, K.K., Shaywitz, B.A., & Fletcher, J.M. Developmental lag versus deficit models of reading disability: A longitudinal, individual growth curves analysis. Journal of Educational Psychology, 1996, 88, 3-17.
- Harter, S. (1982). The Perceived Competence Scale for Children. Child Development, 53, 87-97.
- Henderson, R.W., Bergan, J.R., & Hurt, M. (1972). Development and validation of the Henderson Environmental Learning Process Scale. Journal of Social Psychology, 88, 185-196.
- Hoover, W.A., & Gough, P.B. (1990). The simple view of reading. Reading and Writing, 2, 127-60.
- Juel, C. (1988). Learning to read and write: A longitudinal study of 54 children from first through fourth grades. Journal of Educational Psychology, 80, 437-447.
- Kaufman, A.S., & Kaufman, N.L. (1985). Kaufman Test of Educational Achievement. Circle Pines, MN: American Guidance Service.
- Lyon, R. G. (1995). Towards a definition of dyslexia. Annals of Dyslexia, 45, 3-27.
- Mann, V. A., & Liberman, I. Y. (1984). Phonological awareness and verbal short-term memory. Journal of Learning Disabilities, 17, 592-599.
- McBride-Chang, C., & Manis, F.R. (1996). Structural invariance in the associations of naming speed, phonological awareness, and verbal reasoning in good and poor readers: A test of the double deficit hypothesis. Reading and Writing: An Interdisciplinary Journal, 8, 323-339.
- Pearson, P.D., & Fielding, L. (1994). Comprehension instruction. In R. Barr, M.L. Kamil, P. Mosenthal, & P.D. Pearson (Eds.), Handbook of reading research (Vol. I); pp. 815-860. White Plains, NY: Longman

- Perfetti, C. A. (1985). Reading ability. New York: Oxford University Press.
- Pressley, M. (1998). Reading instruction that works: The case for balanced teaching. New York: Guilford.
- Rosner, J., & Simon, D. P. (1971). The Auditory Analysis Test: An initial report. Journal of Learning Disabilities, 4, 384-392.
- Satz, P., Taylor, H. G., Friel, J., & Fletcher, J. M. (1978). Some developmental and predictive precursors of reading disabilities: A six-year follow-up. In A. L. Benton & D. Pearl (Eds.), Dyslexia: An appraisal of current knowledge (pp. 457-501). New York: Oxford.
- Satz, P., & Fletcher, J.M. (1982). Florida Kindergarten Screening Battery. Florida: Psychological Assessment Resources.
- Schatschneider, C., Francis, D.J., Foorman, B.R., Fletcher, J.M., & Mehta, P. (1999). The dimensionality of phonological awareness: An application of item response theory. Journal of Educational Psychology.
- Share, D. L., Jorm, A. F., MacLean, R., & Matthews, R. (1984). Sources of individual differences in reading acquisition. Journal of Educational Psychology, 76, 1309-1324.
- Shaywitz, S.E., Holahan, J.M., Marchione, K.E., Sadler, A.E., & Shaywitz, B.A. (1992). The Yale Children's Inventory. In S.E. Shaywitz & B.A. Shaywitz (Eds.), Attention deficit disorder come of age (pp. 29-68). Austin, TX: Pro-Ed.
- Shaywitz, B. A., Fletcher, J. M., Holahan, J. M., and Shaywitz, S. E. (1992). Discrepancy compared to low achievement definitions of reading disability: Results from the Connecticut Longitudinal Study. Journal of Learning Disabilities, 25, 639-648.
- Stanovich, K. E. (1991). Discrepancy definitions of reading ability: Has intelligence led us astray? Reading Research Quarterly, 26, 7-29.
- Stanovich, K. E., & Siegel, L. S. (1994). Phenotypic performance profile of children with reading disabilities: A regression-based test of phonological-core variable-difference model. Journal of Educational Psychology, 86, 24-53.
- Stuart, M. (1995). Prediction and qualitative assessment of five-and six-year-old children's reading: A longitudinal study. British Journal of Educational Psychology, 65, 287,296.
- Torgesen, J. K., Wagner, R.K., & Rashotte, C. A. (1994). Longitudinal studies of phonological processing and reading. Journal of Learning Disabilities, 27, 276-286.
- Torgesen, J.K., Wagner, R., & Raschotte, C.A. (in press). Comprehensive Assessment of Phonological Processes. Austin, TX: Pro-Ed.

- Torgesen, J. K., Wagner, R.K., Rashotte, C. A., Alexander, A. W., & Conway, T. (1997). Preventive and remedial interventions for children with severe reading disabilities. Learning Disabilities: A Multi-Disciplinary Journal, 8, 51-61.
- Treiman, R., Tincoff, R., Rodriguez, K., Simon-Mouzaki, A., & Francis, D.J. (in press). The foundations of literacy: Learning the sounds of letters. Child Development.
- Vellutino, F.R., Scanlon, D.M., Sipay, E., Small, S., Pratt, A., Chen, R., & Denckla, M. (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 basic causes of specific reading disability. Journal of Educational Psychology 88, 601-638.
- Wagner, R.K., & Torgesen, J.K. (1987). The nature of phonological processing and its causal role in the acquisition of reading skills. Psychological Bulletin, 101, 192-212.
- Wagner, R., Torgesen, J.K., & Rashotte, C.A. (1994). Development of reading-related phonological processing abilities: New evidence of bi-directional causality from a latent variable longitudinal study. Developmental Psychology, 30, 73-87.
- Wechsler, D. (1974). Wechsler Intelligence Scale for Children-Revised. San Antonio, TX: Psychological Corporation.
- Wiederholt, L.J. (1986). Formal Reading Inventory. Austin, TX: Pro-Ed.
- Wiederholt, L.J. (1994). Gray Oral Reading Test-3. Austin, TX: Pro-Ed.
- Wolf, M., Bally, H., Morris, R. (1986). Automaticity, retrieval processes, and reading: A longitudinal study in average and impaired readers. Child Development, 57, 988-1000.
- Wolf, M., Pfeil, C., Lotz, R., & Biddle, K. (1994). Towards a more universal understanding of the developmental dyslexias: The contribution of orthographic factors. In V.W. Berninger, The varieties of orthographic knowledge (Vol. 1, pp. 137-171). Dordrecht, The Netherlands: Kluwer Academic Publishers.
- Woodcock, R.W., & Johnson, M.B. (1989). Woodcock-Johnson Psycho-Educational Battery-Revised. Allen, TX: DLM Teaching Resources.