TECHNICAL REPORT

TPRI (2010-2014 EDITION)

Children's Learning Institute

University of Texas-Houston Health Science Center

Texas Institute for Measurement, Evaluation, and Statistics

University of Houston

Introduction

The Texas Primary Reading Inventory (TPRI) is a teacher-administered assessment of reading skills for children in kindergarten, Grade 1, Grade 2 and Grade 3. It was designed to comply with the requirements of Texas Education Code 28.006 by providing a research-based assessment of early reading skills, which is required for all children in Kindergarten-Grade 2 attending public school in Texas. The primary purposes of the TPRI are to facilitate a teacher's capacity to a) identify children at-risk for reading difficulties, including dyslexia, in Grades K-2; and b) set learning objectives and develop instructional plans 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, now the Children's Learning Institute (CLI)) at The University of Texas-Houston Health Science Center and the Texas Institute for Measurement, Evaluation, and Statistics (TIMES) at the University of Houston were contracted to revise the TPRI in order to ensure alignment with a) the Texas Essential Knowledge and Skills (TEKS); and b) research on reading skills development. In addition, CLI/TIMES were asked to provide evaluations of the reliability, validity, and implementation of the TPRI, which is an ongoing process.

A description, rationale and purposes of the TPRI can be found in the teachers' guide. The TPRI is designed for administration at the beginning and end of kindergarten, Grade 1, Grade 2 and Grade 3. Assessments are also possible at the midyear to monitor progress, but the forms for these assessments are essentially the same at the beginning of the year.

At beginning and end of kindergarten, beginning and end of Grade 1, beginning of Grade 2, and beginning of Grade 3 the TPRI consists of both 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 the additional assessment portions of the inventory. This saves teachers considerable time because they do not need to administer the entire inventory to these low-risk children. 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 and

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.

In 1998, CARS/TIMES completed an initial assessment of the reliability, validity, and teacher responses to the 1997-1998 edition of the TPRI. Overall, the reliability and validity of the TPRI were satisfactory and teacher response was positive. However, there were components of the 1997-1998 version of the TPRI where the reliability was not adequate, particularly tasks involving book and print awareness, and some comprehension tasks. Teachers provided many comments that suggested ways in which the TPRI could be improved, especially in formatting, directions, and scoring. The 1998 study was conducted on a relatively small sample in Houston. Thus, it was necessary to evaluate the TPRI in a larger, more diverse sample, where data was collected in school districts across the state.

To address these issues, the 1998 TPRI was revised to create the 1998-1999 edition of the TPRI. The most significant revisions involved the development of standardized scoring rubrics for the passages used to assess listening and reading comprehension skills, a major concern of teachers. Many items were re-written to improve the reliability of specific tasks of the TPRI. Directions and formats were changed to make the TPRI easier to administer. An instructional activities guide was added to help teachers develop instructional plans.

After revision, the psychometric characteristics and implementation of the 1998-1999 edition of the TPRI were again evaluated. A statewide study was conducted by CARS and TIMES. Given the research underlying the development of the TPRI and the analysis of validity completed in the first study (see 1997-1998 Technical Report), validity was not viewed as a major issue. Rather, the issues of primary importance involved the reliability of the 1998-1999 TPRI, possible ethnic or gender bias, and its implementation in the field by schools and teachers. Thus, a study was designed that would permit the collection of TPRIs actually administered by teachers in kindergarten, Grade 1, and Grade 2, along with teachers' evaluations of the TPRI. This study found that with the exception of tasks involving Book and Print Awareness and listening/reading comprehension, the overall reliability of the1998-1999 TPRI tasks met commonly accepted standards for reliability. There was no evidence for significant item bias by

ethnicity or gender. The response of teachers in the urban, small city, and rural districts was neutral to positive. The suburban district was neutral to negative.

These results indicated that further revisions of the TPRI were called for. In the 1999-2000 edition, the Book and Print Awareness tasks were changed to "warm-up" exercises, as reliability was consistently low. In 2004, a TPRI – Grade 3 edition was created with similar structure and scoring as Grade 2. In the 1999-2000 edition, the 2002-2003 edition and again in the 2004-2006 and 2006-2010 editions, some specific items were revised to improve the reliability on some tasks and additional authentic stories were added. These revisions did not represent major administration changes. In the 1999-2000 and 2004-2006 editions major formatting changes were made in response to teacher feedback to make the TPRI easier to administer. Additional training guides and materials on intervention strategies have been regularly updated in response to suggestions from teachers and others who have interacted with CARS/TIMES in the development of the TPRI.

2008-2009 TPRI Development Study

Unlike the smaller-scale development studies prior to the more recent editions, the 2008-2009 development study was designed to assess and when possible improve the validity and overall reliability of the entire TPRI. Each of the screens was re-evaluated to ensure the maintenance of predictive validity. Also for the reading portion (story forms), stories were rewritten and comprehension items reconstructed to be shorter and equated. Equipercentile equating techniques were used to reduce form bias as well as to provide teachers with a means to make appropriate gain comparisons on oral reading fluency across timepoints. We conducted the study in school districts in and outside of Houston, TX. Schools in these districts were invited to participate. In the consenting schools, students and parents of students had the opportunity to decline participation. Student assessment began with the child's assent and was terminated upon either completion of the assessment or if the student indicated to the examiner the desire to finish. Examiners were trained on student assent procedures in accordance with APA guidelines.

As Table 1 shows, 3821 children from 203 classrooms in 16 schools participated in the 2008-2009 TPRI Development Study. The sample was roughly comparable in gender representation (49% male, 51% female) and ethnically diverse, including over 1136 students who were African-American, over 1178 who were Hispanic, and over 1244 who were identified as White. The schools' provided gender and ethnicity data, which were missing or not available for about 20% of the students, accounting for the discrepancies in the total sample size relative to the gender and ethnicity analyses. The sample sizes were adequate to estimate sources of bias for each item of the TPRI, a major goal of the development study. Revalidation of the screening section for Grade 3 was not included in this study as it had been last validated more recently, during the 2004-2006 revision. In each school, research assistants administered new or current sections of the TPRI to students using scannable data collection forms developed by TIMES.

The development study took place during the 2008-2009 academic year. The results are organized into three sections: 1) reliability of the TPRI; 2) evaluation of item bias by ethnicity and gender; and 3) predictive validity of the screens. Although we report reliability data by gender and ethnicity, in some instances the sample size was barely adequate to obtain good estimates of reliability in these subtests of the sample. These data are reported to see if any patterns of differential reliability emerge.

Reliability of the TPRI

There are twelve forms of the TPRI: beginning, middle, and end of Kindergarten, Grade 1, Grade 2, and Grade 3; however, there are only four different forms for the optional inventory tasks. That is, one form to use throughout Kindergarten, one form to use throughout Grade 1, one for Grade 2, and one for Grade 3. The teacher's guide provides explanations for each of these forms and the rationale underlying their development. For each subtest and also for each task within a subtest, Cronbach's alpha was computed. The alpha coefficient ranges from 0 to 1.0 and was reported as an index of internal consistency. High alpha coefficients indicate that item variability is small relative to total test variability, or that all items perform similarly, and were all measuring the same construct. We evaluated the

practical significance of the reliability coefficients as follows: Poor (0-.39), Adequate (.40-.59), Good

(.60-.79) and Excellent (.80-1.0). These estimates of practical significance are arbitrary, but

conventional, and provide a useful heuristic for interpreting the reliability data. As the screens are used

to make decisions about individual children, we required coefficients at least in the "good" or "excellent"

range. Since there were fewer items and smaller samples for 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 .60

("good") as acceptable for alpha coefficients based on the entire sample, and .40 for alpha coefficients

based on ethnicity or gender. This is in part due to the fact that Cronbach's alpha is a lower-bound

estimate of true reliability, and can be strongly affected by restrictions in range of the sample from which

it is estimated.

Reliability of Kindergarten Forms

Table 2 shows the reliability for the two screens (beginning of year and end of year), 10 inventory tasks, one warm-up task, and one optional task (Word Reading task, optional at the end of year) for the kindergarten TPRI collapsing across ethnicity and gender. The screens for both the two beginning and two end of the year screening tasks have high alpha coefficients in the upper part of the excellent range. The alpha for the entire set of Phonological Awareness items is .91. Alphas for the 5 tasks individually range from .73 to .9. The alpha for the entire set of Graphophonemic Knowledge items is .93; alphas for the tasks range from .84 to .93. The alpha for the optional Word Reading Task is .92. Overall, alphas for 15 of the 16 screening and inventory tasks are above .7 (i.e., good); indeed, alphas for 10 of the 16 tasks are above .8 (i.e., excellent). The alpha below .6 involves the print awareness task. As reported in the 1997-1998 and 1998-1999 Technical Reports, the Book and Print Awareness task had inadequate reliability and was used in this version only as "warm-up" tasks. Following the same recommendation, the Book and Print Awareness task is again an optional "warm-up" task in the 2010 version. The reliability of the three listening comprehension kindergarten stories all hover around .76, considered in the "good" range. These estimates are higher than in prior versions.

In Table 3, alphas are computed separately by ethnicity for the kindergarten forms. The results are similar to the overall estimates in Table 2, showing excellent reliability for the screens, and good to excellent reliability for all inventory tasks except for Book and Print Awareness. Table 4 reports the kindergarten reliabilities by gender. There is no evidence for major gender differences in the reliabilities across each task and the patterns parallel those apparent in the overall results.

Reliability of Grade 1 Forms

Table 5 reports the overall reliability results for the Beginning of Grade 1 and End of Grade 1 TPRI tasks, while Tables 6-7 break down the reliabilities by ethnicity and gender, respectively. In Table 5, we see that the four screening tasks have excellent reliabilities (.88-.92). The alphas for the entire sets of items for the Phonological Awareness, Graphophonemic Knowledge, and Word Reading subtests are .81, .88, and .94, respectively. The inventory subtest tasks range from .62 to .84. The Reading

Comprehension forms hover around .71 (range .70 to .73). Of the 20 Grade 1 tasks, 16 have alphas higher than .70 and all are greater than .6. The reliabilities for the stories are higher than in prior editions.

Tables 6 and 7 do not reveal major differences in reliability estimates by ethnicity and gender for the Grade 1 TPRI tasks (Table 6). All are in the Good to Excellent range, except for Final Consonant Substitution within the Black and Hispanic ethnicities, .59 and .59; however, the reliability for students of White ethnicity is .65. This suggests that the lower reliabilities for this one task are task dependent and not a sign of particular bias for ethnicity. Ethnic and gender bias will be specifically evaluated in the next section. Other than this task, no Grade 1 tasks show lower than good reliability and there is no specific pattern of bias apparent for either ethnicity or gender.

Reliability of Grade 2 Forms

Table 8 reports the reliabilities for the overall Grade 2 sample, while Tables 9-10 break down the results by ethnicity and gender, respectively. In Table 8, the Grade 2 form shows excellent reliability for the Word Reading screening task. Also, the Graphophonemic Knowledge (Spelling) and Word Reading portions of the inventory are excellent, .87 and .89, respectively. The reliabilities for the subsets of items are in the good to excellent range. Unlike the phonological awareness and graphophonemic knowledge inventory tasks in earlier grades, the Spelling and Word Reading sets are only different "sets" in practical application. Consider the difference between a rhyming task and an elision task asking the student to delete a final sound from a word. These both represent phonological awareness skills but are also distinctly different tasks. For this reason, the reliability for the full scale and individual tasks are presented. The procedures and scoring for the Word Reading and Spelling Tasks are identical, so it is appropriate to primarily interpret the full-scale reliability coefficient (excellent). Nonetheless, we also report the subscale sets of items. For both Graphophonemic Knowledge and Word Reading, the reliabilities of all sets are in the "good" range, ranging from .60 to .75. The reading comprehension stories range from .67 to .68. As with prior versions, although these are in the acceptable "good" range, higher reliabilities here are desirable. Tables 9 and 10 show generally similar reliability estimates across

ethnicity and gender which is an improvement over prior versions and no particular pattern is apparent

that would indicate systematic bias.

Reliability of Grade 3 Forms

Table 11 reports the reliabilities for the overall Grade 3 sample, while Tables 12-13 break down

the results by ethnicity and gender, respectively. In Table 11, the Grade 3 form shows excellent

reliability for the Word Reading screening task, alpha = .88. Also, the Spelling and Word Reading

portions of the inventory are excellent, .90 and .89 respectively. The reliabilities for the subsets of items

are in the good to excellent range. The procedures and scoring for the Word Reading and Spelling subsets

are identical, so it is appropriate to primarily interpret the full-scale reliability coefficient (excellent).

Nonetheless, we also report the subscale sets of items. For both Spelling and Word Reading, the

reliabilities of all sets are in the "good" range, ranging from .68 to .74. The reading comprehension

stories range from .69 to .71. As with prior versions, although these are in the acceptable "good" range,

higher reliabilities here are desirable. Tables 12 and 13 show generally similar reliability estimates

across ethnicities and genders.

Summary: Reliability

The reliability analyses consistently show excellent reliabilities for all of the screening tasks. All

but one TPRI tasks are in the good to excellent range at each grade level, the lone exception being Book

and Print Awareness in Kindergarten. For this reason, it is an optional task. Additionally, 24 out of the 31

(over 77%) inventory tasks show reliability coefficients greater than .70. As we reported in prior

technical reports, the Book and Print Awareness task has inadequate reliability. Still, there are some

reliability estimates in the "good" range and these should be considered for improvement in future

versions. There are no apparent systematic differences in reliability by ethnicity and gender, which tend

to be consistent. The inconsistencies are most likely related to the small number of items and low sample

sizes for some of the gender and ethnicity analyses. While reporting reliability by gender and ethnicity is

important for identifying patterns of less than adequate reliability, the key question is whether specific

items function differently by virtue of ethnicity and gender. This is addressed in the next section.

Differential Item Functioning Analysis

Item response theory (IRT) models were employed for detecting differential item bias. IRT permits comparisons of item functioning between groups in terms of the probability that performance of that item for each group is different at the same level of ability. To conduct these analyses, an IRT model is constructed that estimates item parameters for each group of interest (e.g., ethnicity), and compares the parameters obtained for this model to a model in which group membership is ignored. If the models are not different, this indicates that the differences between groups on an item are best explained solely by ability and that group membership does not contribute to differential performance of an item. This would indicate that the item is not biased. It is important to recognize that some items will show evidence for differential item functioning (DIF) solely by chance. The goal is to keep the total number of items indicating DIF below 5%.

Tables 14-16 summarize the number of items showing DIF by gender and ethnicity (White/Black, White/Hispanic). Of the 130 items on the Kindergarten form of the TPRI, four (3%) showed gender differences. Seven (5%) showed differences between Blacks and Whites. However, three of these favored Whites and four favored Blacks. Different subtests showed DIF relative to prior versions. There was 1 item (<1%) that showed DIF for the White/Hispanic analysis. The overall rate was at or below 5.4% for each analysis.

Table 15 shows only four of 137 items demonstrated DIF between genders on the Grade 1 form, three favoring boys. In the White/Black comparison, 6 items showed DIF, 4 of them favoring White students. For the White/Hispanic comparison, there were 9 items showing DIF, but 5 of these were advantageous to Hispanic students. Accordingly, there does not appear to be a consistent pattern of bias. The overall rates of DIF for each analysis were below 5% for the gender and White/Black comparison, and 6.6% for the White/Hispanic comparison.

A similar pattern is apparent for Grade 2 (Table 16). For the 88 items, 1 item shows DIF by gender, 3 for the White/Black comparison, with 2 favoring Blacks, and 3 for the White/Hispanic

comparison, with 1 favoring Hispanics. The overall rates are 3.5% or less for each of these analyses and there is no pattern of bias that would be of significant concern.

Grade 3 sample sizes for subgroups were not large enough to appropriately model DIF. While these analyses have been excluded from this report, there is little evidence to support systematic bias in any of the items from earlier grades and a similar conclusion is not unreasonable for the third grade items.

Summary: Item Bias

There is no evidence for systematic item bias by virtue of ethnicity or gender for any of the forms of the TPRI. The overall rates of DIF for any specific comparison are almost uniformly below 5%, the lone exception being the first grade White/Hispanic comparison. However, for those 9 items with DIF, 5 favored the Hispanic students and 4 favored White students. Also, the items affected tend to be on different tasks relative to prior versions, supporting the absence of systematic bias by item or task.

<center>Predictive Validity of TPRI Screens</center>

Kindergarten Screens

The 2010 Kindergarten screens were revised based upon the existing TPRI screens as well as testing other measures for predictive utility. Performance at the beginning of the year on various screening tasks was compared against outcome measures administered in the late spring, or end of year timeframe. We also screen students at 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. Predictors include measures of letter names, letter sounds, and phonological awareness tasks.

The first step in establishing the best set of predictors involved an examination of all possible combinations in predicting outcomes at the end of the year. To this, 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. Variables were selected if they 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 cutpoint from the equation expressing the relationship of the predictors and outcomes was established. This cut-point was achieved by deliberately and manually adjusting the equation to establish the lowest possible false positive error rate, while also keeping false negative error rates below 10%. Cut-offs that produced the most desirable classification were selected. Tables 17 and 18 summarize the identification tables for the Kindergarten beginning of year and end of year screens relative to end of year outcomes using the measures of letter-sound naming and phonological awareness.

Table 17 shows that only 6 out of the 743 students who were assessed at both BOY and EOY were misclassified at BOY as not at-risk. This represents a false-negative, a more egregious type of error in an educational setting. A false-positive misclassification represents students who were misidentified at BOY as being at-risk but who were not considered at-risk at end of year based upon the end of year outcome measures. This error is less egregious because it merely amounts some additional assessment (i.e., gathering data from the inventory tasks). False-negative errors, on the other hand, should be minimized in order to prevent failure-to-identify students who do show signs of struggle at the end of the year. False-negative rate is determined by the number of misclassifications (6) divided by the total number of students who were at risk on the outcome measure (94), or 6%. Table 18 shows that only 8 out of the 744 students who were assessed using both TPRI and outcome measures at EOY were misclassified as not at-risk. The false-negative rate at end of year was 9%.

Grade 1 Screens

The 2010 Grade 1 screens were similarly revised based upon the existing TPRI screens as well as piloting new items and measures. We employed the same logic and procedures as outlined above for the Kindergarten screens. The First Grade screen at beginning of year consists of three short tasks, a ten item letter-sound identification task, an eight item word reading task, and a 6 item blending phoneme task. The 10 item letter-sound identification task does not factor into the overall decision rule for a student to

be considered "Developed" or "Still Developing" in the screen. It is to provide some carry-over from Kindergarten or to provide first grade teachers some information about their students' letter-sound identification abilities at the beginning of year. To be "Developed" students must correctly read 4 out of the 8 words on the word reading task or blend 5 out of the 6 words on the blending phonemes task. As shown in Table 19, at the end of first grade 129 students were below the 20th percentile on the Woodcock-Johnson Broad Reading cluster. There were 602 students that scored above the 20th percentile. If we had applied the decision rule from data that was collected in the Fall, we would have correctly identified 120 out of these 129 students who ended up below the outcome criterion and 426 out of the 602 who ended up above the criterion. These decision criteria would have incorrectly identified 176 out of the 602 who ended up above the outcome criterion as "Still Developing" (false-positive identification rate); these decision criteria would have incorrectly not-identified 9 out of the 129 who fell below the outcome criterion (false-negative identification rate). The 7% false-negative rate is strong and comparable to prior TPRI screens. In our sample, the revised TPRI screens would have only "missed" 9 out of 731 students.

The screen at the end of first grade performs similarly. The EOY screen is a 12-item word reading task. Students who correctly read 8 out of the 12 items are considered "Developed" and students who correctly read 7 or fewer are considered "Still Developing" on the screen. For the 735 students that we had complete data, 608 were above the 20th percentile on the WJ Broad Reading and 127 were below. The end of year screen for first grade correctly identified 466 and misidentified 142 (false-positive) of the 608 students above threshold. The end of year screen for first grade correctly identified 117 and failed to identify 10 (false-negative) of the 127 students below threshold. Again, the 8% false-negative rate is strong and comparable to prior TPRI screens. In our sample, the revised TPRI screens would have only "missed" 10 out of 735 students with a 12-item word reading task that takes less than 3 minutes to administer.

Grade 2 Screen

The 2010 Grade 2 screen was similarly revised based upon the existing TPRI second grade screen as well as piloting new items and measures. We employed the same logic and procedures as

outlined above for the Kindergarten and first grade screens. The Grade 2 TPRI has a screen at beginning of second grade which consists of 12-item word reading task. Students who correctly read 9 out of the 12 items are considered "Developed" and students who correctly read 8 or fewer are considered "Still Developing" on the screen. For the 814 students that we had complete data, 727 were above the 20th percentile on the WJ Broad Reading and 87 were below. The beginning of year screen for second grade correctly identified 559 and misidentified 168 (false-positive) of the 727 students above threshold. The beginning of year screen for second grade correctly identified 77 and failed to identify 10 (false-negative) of the 87 students below threshold. The 11% false-negative rate is strong and comparable to prior TPRI screens. In our sample, the revised second grade TPRI screen would have only "missed" 10 out of 814 students with a 12-item word reading task that takes less than 3 minutes to administer.

Grade 3 Screen

The 2010 Grade 3 screen was not revised in this edition. The Grade 3 TPRI has a screen at the beginning of the year which consists of a 20-item word reading task. Students who correctly read 19 out of the 20 items are considered "Developed" and students who correctly read 18 or fewer are considered "Still Developing" on the screen. For the 739 students for whom we had complete data, 691 were above the 20th percentile on the WJ Broad Reading and 48 were below. The beginning of year screen for third grade correctly identified 494 and misidentified 197 (false-positive) of the 691 students above threshold. The beginning of year screen for third grade correctly identified 45 and failed to identify 3(false-negative) of the 48 students below threshold. The 6% false-negative rate is strong. In our sample, the third grade TPRI screen would have only "missed" 3 out of 739 students with a 20-item word reading task that takes less than 3 minutes to administer.

Summary: TPRI Screens

The 2010 TPRI Development Study set out to revalidate the existing kindergarten, first, and second grade TPRI screens. To do this, measures were collected in both the fall and spring at every grade level. Outcome measures were administered at the end of the year and students were evaluated against whether they fell above or below a set threshold on the outcome. Balancing correct identification and minimizing false-negative identifications is the primary importance of the TPRI screens. Across the

four grades and six screen forms (K-BOY, K-EOY, G1-BOY, G1-EOY, G2-BOY,G3-BOY), we evaluated the revised screens with data from 4581 student outcomes. While the TPRI screens would have correctly identified over 70% of those students, it is more instructive to consider that the TPRI screens would have failed to identify only 46 out of 4506 (~1%) students through the use of short assessments that take less than 3 to 5 minutes per student.

Conclusions

In general, the 2010 version of the TPRI shows strong reliability with little evidence of gender and ethnic bias. In most cases, the reliability has improved from previous versions. Also, from a reliability viewpoint, the Listening and Reading Comprehension stories were improved through the development study; however they might be improved further. For this revision, the development study set out to re-validate the screens for their predictive validity focusing on minimizing false-negative identifications. The implications for false-negative identifications are of substantially more concern than those for false-positive identifications. The TPRI screens show excellent predictive validity with low false-negative rates. In conclusion and similar to prior versions, the 2010 version of the TPRI demonstrates high validity and high reliability.

Table 1. Description of sample for the 2008-2009 TPRI Development Study.

Entity	Grade Level				
	Kindergarten	Grade 1	Grade 2	Grade 3	Total
Schools	16	16	16	11	16
Classrooms	57	55	57	34	203
Students	1060	1056	1118	587	3821

Table 2. Overall reliabilities for Kindergarten tasks.

Total Alphas		
Subtest	N	Alpha
Screen – BOY (Full scale 18 items)	743	0.90
Screen 1: Letter Sound: Letter Sound (10 items)	*743*	*0.85*
Screen 2: Blending Onset-Phonemes (8 items)	*743*	*0.88*
Screen – EOY (Full scale 18 items)	743	0.86
Screen 3: Letter Sound: Letter Sound (10 items)	*743*	*0.85*
Screen 4: Blending Onset-Rhymes and Phonemes (8 items)	*743*	*0.91*
Phonological Awareness (Full scale 25 items)	686	0.91
Rhyming Task (5 items)	*687*	*0.80*
Blending Word Parts Task (5 items)	*686*	*0.77*
Blending Phonemes Task (5 items)	*689*	*0.73*
Deleting Initial Sounds Task (5 items)	*689*	*0.90*
Deleting Final Sounds Task (5 items)	*687*	*0.87*
Graphophonemic Knowledge (Full scale 36 items)	501	0.93
Letter Name Identification Task (26 items)	*501*	*0.93*
Letter to Sound Linking Task (10 items)	*688*	*0.84*
Listening Comprehension		
Listening Comprehension BOY (6 items)	*224*	*0.76*
Listening Comprehension MOY (6 items)	*221*	*0.76*
Listening Comprehension EOY (6 items)	*230*	*0.76*
Word Reading (Optional Task, 10 items)	689	0.92
Book and Print Awareness (Optional Warm-up, 5 items)	611	0.36

BOY = Beginning of Year; MOY = Middle of Year; EOY = End of Year

Table 3. Reliability by Ethnicity for Kindergarten tasks.

Alphas by Ethnicity						
Subtest	N	Black	N	Hispanic	N	White
Screen – BOY (Full scale 18 items)	184	0.88	162	0.89	198	0.92
Screen 1: Letter Sound: Letter Sound (10 items)	*184*	*0.87*	*162*	*0.82*	*198*	*0.91*
Screen 2: Blending Onset-Phonemes (8 items)	*184*	*0.92*	*162*	*0.86*	*198*	*0.88*
Screen – EOY (Full scale 18 items)	184	0.86	162	0.87	198	0.84
Screen 3: Letter Sound: Letter Sound (10 items)	*184*	*0.82*	*162*	*0.86*	*198*	*0.89*
Screen 4: Blending Onset-Rhymes and Phonemes (8 items)	*184*	*0.95*	*162*	*0.94*	*198*	*0.92*
Phonological Awareness (Full scale 25 items)	163	0.92	132	0.89	181	0.91
Rhyming Task (5 items)	*163*	*0.67*	*132*	*0.77*	*181*	*0.62*
Blending Word Parts Task (5 items)	*163*	*0.82*	*132*	*0.75*	*181*	*0.71*
Blending Phonemes Task (5 items)	*163*	*0.82*	*132*	*0.71*	*181*	*0.65*
Deleting Initial Sounds Task (5 items)	*163*	*0.91*	*132*	*0.88*	*181*	*0.91*
Deleting Final Sounds Task (5 items)	*163*	*0.87*	*132*	*0.84*	*181*	*0.88*
Graphophonemic Knowledge (Full scale 36 items)	163	0.95	132	0.93	181	0.87
Letter Name Identification Task (26 items)	*163*	*0.95*	*132*	*0.94*	*181*	*0.86*
Letter to Sound Linking Task (10 items)	*163*	*0.89*	*132*	*0.87*	*181*	*0.75*
Listening Comprehension						
Listening Comprehension BOY (6 items)	*75*	*0.75*	*65*	*0.77*	*84*	*0.72*
Listening Comprehension MOY (6 items)	*75*	*0.76*	*65*	*0.76*	*84*	*0.76*
Listening Comprehension EOY (6 items)	*75*	*0.73*	*65*	*0.76*	*84*	*0.74*
Word Reading (Optional Task, 10 items)	163	0.93	134	0.91	182	0.93
Book and Print Awareness (Optional Warm-up, 5 items)	84	0.47	261	0.35	111	0.32

BOY = Beginning of Year; MOY = Middle of Year; EOY = End of Year

Table 4. Reliabilities by Gender for Kindergarten tasks.

Alphas by Gender				
Subtest	N	Male	N	Female
Screen – BOY (Full scale 18 items)	327	0.9	341	0.86
Screen 1: Letter Sound: Letter Sound (10 items)	*327*	*0.81*	*341*	*0.85*
Screen 2: Blending Onset-Phonemes (8 items)	*327*	*0.91*	*341*	*0.92*
Screen – EOY (Full scale 18 items)	327	0.82	341	0.87
Screen 3: Letter Sound: Letter Sound (10 items)	*327*	*0.87*	*341*	*0.81*
Screen 4: Blending Onset-Rhymes and Phonemes (8 items)	*327*	*0.94*	*341*	*0.88*
Phonological Awareness (Full scale 25 items)	299	0.89	312	0.87
Rhyming Task (5 items)	*299*	*0.82*	*312*	*0.77*
Blending Word Parts Task (5 items)	*299*	*0.73*	*312*	*0.76*
Blending Phonemes Task (5 items)	*299*	*0.7*	*313*	*0.73*
Deleting Initial Sounds Task (5 items)	*299*	*0.94*	*313*	*0.92*
Deleting Final Sounds Task (5 items)	*299*	*0.86*	*312*	*0.84*
Graphophonemic Knowledge (Full scale 36 items)	209	0.94	217	0.95
Letter Name Identification Task (26 items)	*209*	*0.9*	*217*	*0.91*
Letter to Sound Linking Task (10 items)	*209*	*0.82*	*313*	*0.82*
Listening Comprehension				
Listening Comprehension BOY (6 items)	*110*	*0.76*	*114*	*0.79*
Listening Comprehension MOY (6 items)	*108*	*0.76*	*113*	*0.72*
Listening Comprehension EOY (6 items)	*113*	*0.78*	*117*	*0.72*
Word Reading (Optional Task, 10 items)	301	0.94	313	0.96
Book and Print Awareness (Optional Warm-up, 5 items)	273	0.35	237	0.38

MOY = Middle of Year; EOY = End of Year

Table 5. Overall reliabilities for Grade 1 tasks.

Subtest	N	Alpha
Total Alphas		
Screen – BOY (Full scale 24 items)	731	0.89
Screen 1: Letter Sound: Letter Sound (10 items)	*731*	*0.88*
Screen 2: Word Reading (8 items)	*731*	*0.92*
Screen 3: Blending Phonemes (6 items)	*731*	*0.88*
Screen – EOY Word Reading (12 items)	735	0.91
Phonological Awareness (Full scale 20 items)	694	0.81
Blending Word Parts Task (5 items)	*694*	*0.74*
Blending Phonemes Task (5 items)	*694*	*0.65*
Deleting Initial Sounds Task (5 items)	*694*	*0.72*
Deleting Final Sounds Task (5 items)	*694*	*0.84*
Graphophonemic Knowledge (Full scale 25 items)	694	0.88
Initial Consonant Substitution (5 items)	*694*	*0.67*
Final Consonant Substitution (5 items)	*694*	*0.62*
Middle Vowel Substitution (5 items)	*694*	*0.66*
Initial Blending Substitution (5 items)	*694*	*0.81*
Blends in Final Position (5 items)	*694*	*0.82*
Word Reading (Full scale 20 items)	694	0.94
Set 1 (5 items)	*694*	*0.82*
Set 2 (5 items)	*694*	*0.83*
Set 3 (5 items)	*694*	*0.79*
Set 4 (5 items)	*694*	*0.74*
Reading Comprehension		
Reading Comprehension BOY (12 items)	*339*	*0.73*
Reading Comprehension MOY (12 items)	*434*	*0.70*
Reading Comprehension EOY (12 items)	*322*	*0.71*

BOY = Beginning of Year; EOY = End of Year

Table 6. Reliabilities by Ethnicity for Grade 1 tasks.

Subtest	N	Black		Hispanic		White
Alphas by Ethnicity						
Screen – BOY (Full scale 24 items)	165	0.91	153	0.92	184	0.93
Screen 1: Letter Sound: Letter Sound (10 items)	*165*	*0.92*	*153*	*0.88*	*184*	*0.91*
Screen 2: Word Reading (8 items)	*165*	*0.9*	*153*	*0.88*	*184*	*0.96*
Screen 3: Blending Phonemes (6 items)	*165*	*0.88*	*153*	*0.88*	*184*	*0.92*
Screen – EOY Word Reading (12 items)	165	0.95	153	0.93	184	0.88
Phonological Awareness (Full scale 20 items)	165	0.84	153	0.82	184	0.8
Blending Word Parts Task (5 items)	*165*	*0.74*	*153*	*0.75*	*184*	*0.76*
Blending Phonemes Task (5 items)	*165*	*0.68*	*153*	*0.61*	*184*	*0.65*
Deleting Initial Sounds Task (5 items)	*165*	*0.71*	*153*	*0.7*	*184*	*0.68*
Deleting Final Sounds Task (5 items)	*165*	*0.82*	*153*	*0.87*	*184*	*0.85*
Graphophonemic Knowledge (Full scale 25 items)	165	0.92	153	0.86	184	0.91
Initial Consonant Substitution (5 items)	*165*	*0.7*	*153*	*0.66*	*184*	*0.7*
Final Consonant Substitution (5 items)	*165*	*0.59*	*153*	*0.59*	*184*	*0.65*
Middle Vowel Substitution (5 items)	*165*	*0.65*	*153*	*0.66*	*184*	*0.64*
Initial Blending Substitution (5 items)	*165*	*0.8*	*153*	*0.78*	*184*	*0.84*
Blends in Final Position (5 items)	*165*	*0.78*	*153*	*0.8*	*184*	*0.84*
Word Reading (Full scale 20 items)	165	0.91	153	0.95	184	0.98
Set 1 (5 items)	*165*	*0.85*	*153*	*0.84*	*184*	*0.84*
Set 2 (5 items)	*165*	*0.79*	*153*	*0.82*	*184*	*0.84*
Set 3 (5 items)	*165*	*0.75*	*153*	*0.77*	*184*	*0.82*
Set 4 (5 items)	*165*	*0.7*	*153*	*0.77*	*184*	*0.76*
Reading Comprehension						
Reading Comprehension BOY (12 items)	*88*	*0.72*	*82*	*0.69*	*98*	*0.71*
Reading Comprehension MOY (12 items)	*114*	*0.69*	*106*	*0.66*	*127*	*0.73*
Reading Comprehension EOY (12 items)	*84*	*0.75*	*78*	*0.68*	*93*	*0.68*

BOY = Beginning of Year; EOY = End of Year

Table 7. Reliabilities by Gender for Grade 1 tasks.

Alphas by Gender				
Subtest	N	Male		Female
Screen – BOY (Full scale 24 items)	329	0.88	343	0.87
Screen 1: Letter Sound: Letter Sound (10 items)	*329*	*0.89*	*343*	*0.89*
Screen 2: Word Reading (8 items)	*329*	*0.89*	*343*	*0.92*
Screen 3: Blending Phonemes (6 items)	*329*	*0.91*	*343*	*0.87*
Screen – EOY Word Reading (12 items)	329	0.89	343	0.94
Phonological Awareness (Full scale 20 items)	329	0.84	343	0.83
Blending Word Parts Task (5 items)	*329*	*0.73*	*343*	*0.73*
Blending Phonemes Task (5 items)	*329*	*0.61*	*343*	*0.69*
Deleting Initial Sounds Task (5 items)	*329*	*0.76*	*343*	*0.7*
Deleting Final Sounds Task (5 items)	*329*	*0.81*	*343*	*0.82*
Graphophonemic Knowledge (Full scale 25 items)	329	0.91	343	0.87
Initial Consonant Substitution (5 items)	*329*	*0.63*	*343*	*0.7*
Final Consonant Substitution (5 items)	*329*	*0.61*	*343*	*0.61*
Middle Vowel Substitution (5 items)	*329*	*0.63*	*343*	*0.64*
Initial Blending Substitution (5 items)	*329*	*0.8*	*343*	*0.82*
Blends in Final Position (5 items)	*329*	*0.81*	*343*	*0.78*
Word Reading (Full scale 20 items)	329	0.9	343	0.98
Set 1 (5 items)	*329*	*0.84*	*343*	*0.85*
Set 2 (5 items)	*329*	*0.82*	*343*	*0.86*
Set 3 (5 items)	*329*	*0.76*	*343*	*0.83*
Set 4 (5 items)	*329*	*0.73*	*343*	*0.78*
Reading Comprehension				
Reading Comprehension BOY (12 items)	*155*	*0.72*	*162*	*0.71*
Reading Comprehension MOY (12 items)	*202*	*0.68*	*210*	*0.73*
Reading Comprehension EOY (12 items)	*147*	*0.75*	*153*	*0.68*

BOY = Beginning of Year; EOY = End of Year

Table 8. Overall reliabilities for Grade 2 tasks.

Total Alphas		
Subtest	N	Alpha
Screen – BOY Word Reading (12 items)	814	0.88
Graphophonemic Knowledge (Full scale 20 items)	682	0.87
Set 1 (5 items)	*682*	*0.66*
Set 2 (5 items)	*682*	*0.62*
Set 3 (5 items)	*682*	*0.60*
Set 4 (5 items)	*682*	*0.71*
Word Reading (Full scale 20 items)	694	0.89
Set 1 (5 items)	*694*	*0.73*
Set 2 (5 items)	*694*	*0.61*
Set 3 (5 items)	*694*	*0.70*
Set 4 (5 items)	*694*	*0.75*
Reading Comprehension		
Reading Comprehension BOY (12 items)	*470*	*0.68*
Reading Comprehension MOY (12 items)	*382*	*0.68*
Reading Comprehension EOY (12 items)	*370*	*0.67*

BOY = Beginning of Year; EOY = End of Year

Table 9. Reliabilities by ethnicity for Grade 2 tasks.

Subtest	N	Black	N	Hispanic	N	White
Alphas by Ethnicity						
Screen – BOY Word Reading (12 items)	200	0.88	185	0.9	222	0.9
Graphophonemic Knowledge (Full scale 20 items)	164	0.86	152	0.87	182	0.86
Set 1 (5 items)	*164*	*0.69*	*152*	*0.66*	*182*	*0.62*
Set 2 (5 items)	*164*	*0.63*	*152*	*0.65*	*182*	*0.58*
Set 3 (5 items)	*164*	*0.61*	*152*	*0.62*	*182*	*0.61*
Set 4 (5 items)	*164*	*0.75*	*152*	*0.7*	*182*	*0.68*
Word Reading (Full scale 20 items)	167	0.87	155	0.85	186	0.89
Set 1 (5 items)	*167*	*0.72*	*155*	*0.69*	*186*	*0.71*
Set 2 (5 items)	*167*	*0.62*	*155*	*0.64*	*186*	*0.57*
Set 3 (5 items)	*167*	*0.71*	*155*	*0.66*	*186*	*0.66*
Set 4 (5 items)	*167*	*0.73*	*155*	*0.77*	*186*	*0.73*
Reading Comprehension						
Reading Comprehension BOY (12 items)	*107*	*0.71*	*99*	*0.71*	*119*	*0.72*
Reading Comprehension MOY (12 items)	*83*	*0.65*	*77*	*0.7*	*92*	*0.7*
Reading Comprehension EOY (12 items)	*80*	*0.65*	*74*	*0.63*	*89*	*0.65*

BOY = Beginning of Year; EOY = End of Year

Table 10. Reliabilities by gender for Grade 2 tasks.

Alphas by Gender				
Subtest	N	Male	N	Female
Screen – BOY Word Reading (12 items)	388	0.9	404	0.9
Graphophonemic Knowledge (Full scale 20 items)	323	0.84	337	0.9
Set 1 (5 items)	*323*	*0.67*	*337*	*0.7*
Set 2 (5 items)	*323*	*0.62*	*337*	*0.63*
Set 3 (5 items)	*323*	*0.63*	*337*	*0.56*
Set 4 (5 items)	*323*	*0.73*	*337*	*0.7*
Word Reading (Full scale 20 items)	329	0.92	343	0.86
Set 1 (5 items)	*329*	*0.69*	*343*	*0.72*
Set 2 (5 items)	*329*	*0.6*	*343*	*0.58*
Set 3 (5 items)	*329*	*0.69*	*343*	*0.74*
Set 4 (5 items)	*329*	*0.72*	*343*	*0.78*
Reading Comprehension				
Reading Comprehension BOY (12 items)	*220*	*0.68*	*228*	*0.68*
Reading Comprehension MOY (12 items)	*176*	*0.67*	*184*	*0.66*
Reading Comprehension EOY (12 items)	*171*	*0.65*	*177*	*0.63*

BOY = Beginning of Year; EOY = End of Year

Table 11. Overall reliabilities for Grade 3 tasks.

Total Alphas		
Subtest	N	Alpha
Screen – BOY Word Reading (20 items)*	814	0.88
Graphophonemic Knowledge (Full scale 20 items)	423	0.90
Set 1 (5 items)	*423*	*0.69*
Set 2 (5 items)	*423*	*0.71*
Set 3 (5 items)	*423*	*0.72*
Set 4 (5 items)	*423*	*0.68*
Word Reading (Full scale 20 items)	423	0.89
Set 1 (5 items)	*423*	*0.71*
Set 2 (5 items)	*423*	*0.68*
Set 3 (5 items)	*423*	*0.74*
Set 4 (5 items)	*423*	*0.72*
Reading Comprehension		
Reading Comprehension BOY (12 items)	*119*	*0.69*
Reading Comprehension MOY (12 items)	*112*	*0.71*
Reading Comprehension EOY (12 items)	*110*	*0.70*

* The Screen for Grade 3 was not revalidated in 2008-2009 study; data for the Grade 3
   Screen is drawn from the 2004-2006 revision.

BOY = Beginning of Year; EOY = End of Year

Table 12. Reliabilities by ethnicity for Grade 3 tasks.

Alphas by Ethnicity						
Subtest	N	Black	N	Hispanic	N	White
Screen – BOY Word Reading (20 items)	311	0.92	265	0.91	274	0.88
Graphophonemic Knowledge (Full scale 20 items)	191	0.86	232	0.87	164	0.86
Set 1 (5 items)	*191*	*0.74*	*232*	*0.72*	*164*	*0.72*
Set 2 (5 items)	*191*	*0.67*	*232*	*0.72*	*164*	*0.66*
Set 3 (5 items)	*191*	*0.69*	*232*	*0.73*	*164*	*0.65*
Set 4 (5 items)	*191*	*0.70*	*232*	*0.74*	*164*	*0.67*
Word Reading (Full scale 20 items)	167	0.87	155	0.85	186	0.89
Set 1 (5 items)	*167*	*0.72*	*155*	*0.69*	*186*	*0.70*
Set 2 (5 items)	*167*	*0.74*	*155*	*0.66*	*186*	*0.65*
Set 3 (5 items)	*167*	*0.69*	*155*	*0.68*	*186*	*0.70*
Set 4 (5 items)	*167*	*0.71*	*155*	*0.74*	*186*	*0.70*
Reading Comprehension						
Reading Comprehension BOY (12 items)	*35*	*0.67*	*39*	*0.73*	*40*	*0.72*
Reading Comprehension MOY (12 items)	*34*	*0.68*	*38*	*0.70*	*34*	*0.66*
Reading Comprehension EOY (12 items)	*39*	*0.71*	*35*	*0.74*	*39*	*0.72*

* The Screen for Grade 3 was not revalidated in 2008-2009 study; data for the Grade 3 Screen is drawn from the 2004-2006 revision.

BOY = Beginning of Year; EOY = End of Year

Table 13. Reliabilities by gender for Grade 3 tasks.

Alphas by Gender				
Subtest	N	Male	N	Female
Screen – BOY Word Reading (12 items)	410	0.88	404	0.89
Graphophonemic Knowledge (Full scale 20 items)	207	0.85	202	0.87
Set 1 (5 items)	*207*	0.65	*202*	0.67
Set 2 (5 items)	*207*	0.73	*202*	0.72
Set 3 (5 items)	*207*	0.74	*202*	0.72
Set 4 (5 items)	*207*	0.71	*202*	0.65
Word Reading (Full scale 20 items)	207	0.91	202	0.90
Set 1 (5 items)	*207*	0.73	*202*	0.70
Set 2 (5 items)	*207*	0.73	*202*	0.73
Set 3 (5 items)	*207*	0.66	*202*	0.66
Set 4 (5 items)	*207*	0.71	*202*	0.68
Reading Comprehension				
Reading Comprehension BOY (12 items)	*51*	*0.69*	*58*	*0.65*
Reading Comprehension MOY (12 items)	*57*	*0.65*	*52*	*0.63*
Reading Comprehension EOY (12 items)	*56*	*0.66*	*57*	*0.65*

* The Screen for Grade 3 was not revalidated in 2008-2009 study; data for the Grade 3 Screen is drawn from the 2004-2006 revision.

BOY = Beginning of Year; EOY = End of Year

Table 14. Kindergarten TPRI differential item functioning results by task.

SUBTEST	N	Number of items with DIF			
		Gender	White/Black	White/Hisp	Total # Items
Screen - BOY					
Screen 1: Letter Sound: Letter Sound (10 items)	743	0	0	0	10
Screen 2: Blending Onset-Phonemes (8 items)	743	1	0	0	8
Screen - EOY					
Screen 3: Letter Sound: Letter Sound (10 items)	743	0	2	0	10
Screen 4: Blending Onset-Rhymes and Phonemes (8 items)	743	0	0	0	8
Phonological Awareness					
Rhyming Task (5 items)	686	0	1	0	5
Blending Word Parts Task (5 items)	687	0	0	0	5
Blending Phonemes Task (5 items)	686	0	0	0	5
Deleting Initial Sounds Task (5 items)	689	0	1	1	5
Deleting Final Sounds Task (5 items)	689	0	0	0	5
Graphophonemic Knowledge					
Letter Name Identification Task (26 items)	501	1	0	0	26
Letter to Sound Linking Task (10 items)	688	2	0	0	10
Listening Comprehension		0	1	0	
Listening Comprehension BOY (6 items)	224	0	0	0	6
Listening Comprehension MOY (6 items)	221	0	0	0	6
Listening Comprehension EOY (6 items)	230	0	2	0	6
Word Reading	689	0	0	0	10
Book and Print Awareness	228	0	0	0	5
		3.08%	5.38%	0.77%	130

For the 7 items that showed White/Black DIF, 3 of the differences were advantageous to whites. The 1 item that showed White/Hisp DIF was advantageous to whites. For the 4 items that showed gender DIF, 3 of the items were in favor of girls.

Table 15. First Grade TPRI differential item functioning results by task.

		Number of items with DIF			
SUBTEST	N	Gender	White/Black	White/Hisp	Total # Items
Screen – BOY					
Screen 1: Letter Sound: Letter Sound (10 items)	731	0	0	0	10
Screen 2: Word Reading (8 items)	731				8
Screen 3: Blending Phonemes (6 items)	731	0	0	0	6
Screen – EOY Word Reading (12 items)	735	0	0	2	12
Phonological Awareness					
Blending Word Parts Task (5 items)	694	0	0	1	5
Blending Phonemes Task (5 items)	694	0	0	0	5
Deleting Initial Sounds Task (5 items)	694	0	0	0	5
Deleting Final Sounds Task (5 items)	694	1	0	1	5
Graphophonemic Knowledge					
Initial Consonant Substitution (5 items)	694	0	1	0	5
Final Consonant Substitution (5 items)	694	0	1	2	5
Middle Vowel Substitution (5 items)	694	0	0	1	5
Initial Blending Substitution (5 items)	694	2	1	0	5
Blends in Final Position (5 items)	694	0	1	0	5
Word Reading	694				
Set 1 (5 items)	694	0	0	0	5
Set 2 (5 items)	694	0	1	0	5
Set 3 (5 items)	694	1	1	1	5
Set 4 (5 items)	694	0	0	0	5
Reading Comprehension					
Reading Comprehension BOY (12 items)	339	0	0	0	12
Reading Comprehension MOY (12 items)	434	0	0	1	12
Reading Comprehension EOY (12 items)	322	0	0	0	12
		2.92%	4.38%	6.57%	137

For the 6 items that showed White/Black DIF, 4 of them favored whites. For the 9 items that showed White/Hisp DIF, 4 were advantageous to whites. For the 4 items that showed gender DIF, 1 was in favor of girls.

Table 16. Second Grade TPRI differential item functioning results by task.

		Number of items with DIF			
SUBTEST	N	Gender	White/Black	White/Hisp	Total # Items
Screen - BOY Word Reading (12 items)	814	0	0	0	12
Graphophonemic Knowledge					
Set 1 (5 items)	682	1	1	0	5
Set 2 (5 items)	682	0	2	0	5
Set 3 (5 items)	682	0	0	0	5
Set 4 (5 items)	682	0	0	0	5
Word Reading					
Set 1 (5 items)	694	0	0	0	5
Set 2 (5 items)	694	0	0	2	5
Set 3 (5 items)	694	0	0	1	5
Set 4 (5 items)	694	0	0	0	5
Reading Comprehension					
Reading Comprehension BOY (12 items)	470	0	0	0	12
Reading Comprehension MOY (12 items)	382	0	0	0	12
Reading Comprehension EOY (12 items)	370	0	0	0	12
		1.14%	3.41%	3.41%	88

For the 3 items showed White/Black DIF, 1 of the differences was advantageous to Whites. For the 3 items showed White/Hisp. DIF, 2 were advantageous to Whites. The 1 item that showed gender DIF, was in favor of girls.

Table 17. Identification rates for Kindergarten Beginning of Year Screen

OUTCOMES	IDENTIFICATIONS SCREEN 1 & SCREEN 2			CLASSIFICATION METRICS	
WJ-BROAD (END K)	*NO RISK*	*AT RISK*	*TOTAL*	CORRECT CLASSIFY:	61%
				SENSITIVITY:	94%
NO RISK	366	283	649	SPECIFICITY:	56%
AT RISK	6	88	94	Fn RATE:	6%
TOTAL	372	371	743	Fp RATE:	44%

WJ-Broad = the Woodcock-Johnson Broad Reading Cluster score. Risk was determined to be below the 20th percentile on the Broad Reading Cluster.

Table 18. Identification rates for Kindergarten End of Year Screen

OUTCOMES	IDENTIFICATIONS SCREEN 3 & SCREEN 4			CLASSIFICATION METRICS	
WJ-BROAD (END K)	*NO RISK*	*AT RISK*	*TOTAL*	CORRECT CLASSIFY:	65%
				SENSITIVITY:	91%
NO RISK	398	252	650	SPECIFICITY:	61%
AT RISK	8	86	94	Fn RATE:	9%
TOTAL	406	338	744	Fp RATE:	39%

WJ-Broad = the Woodcock-Johnson Broad Reading Cluster score. Risk was determined to be below the 20th percentile on the Broad Reading Cluster.

Table 19. Identification rates for First Grade Beginning of Year Screen

	IDENTIFICATIONS			CLASSIFICATION METRICS	
OUTCOMES	**SCREEN 2 OR SCREEN 3**			CORRECT CLASSIFY:	75%
WJ-BROAD	*NO RISK*	*AT RISK*	*TOTAL*	SENSITIVITY:	93%
NO RISK	426	176	602	SPECIFICITY:	71%
AT RISK	9	120	129	Fn RATE:	7%
TOTAL	435	296	731	Fp RATE:	29%

WJ-Broad = the Woodcock-Johnson Broad Reading Cluster score. Risk was determined to be below the 20th percentile on the Broad Reading Cluster.

Table 20. Identification rates for First Grade End of Year Screen

	IDENTIFICATIONS			CLASSIFICATION METRICS	
OUTCOMES	**SCREEN 4**			CORRECT CLASSIFY:	79%
WJ-BROAD	*NO RISK*	*AT RISK*	*TOTAL*	SENSITIVITY:	92%
NO RISK	466	142	608	SPECIFICITY:	77%
AT RISK	10	117	127	Fn RATE:	8%
TOTAL	476	259	735	Fp RATE:	23%

WJ-Broad = the Woodcock-Johnson Broad Reading Cluster score. Risk was determined to be below the 20th percentile on the Broad Reading Cluster.

Table 21. Identification rates for Second Grade Beginning of Year Screen

OUTCOMES		IDENTIFICATIONS			CLASSIFICATION METRICS	
		SCREEN 1			CORRECT CLASSIFY:	78%
WJ-BROAD		*NO RISK*	*AT RISK*	*TOTAL*	SENSITIVITY:	89%
	NO RISK	559	168	727	SPECIFICITY:	77%
	AT RISK	10	77	87	Fn RATE:	11%
	TOTAL	569	245	814	Fp RATE:	23%

WJ-Broad = the Woodcock-Johnson Broad Reading Cluster score. Risk was determined to be below the 20th percentile on the Broad Reading Cluster.

Table 22. Identification rates for Third Grade Beginning of Year Screen*

OUTCOMES		IDENTIFICATIONS			CLASSIFICATION METRICS	
		SCREEN 1			CORRECT CLASSIFY:	73%
WJ-BROAD		*NO RISK*	*AT RISK*	*TOTAL*	SENSITIVITY:	72%
	NO RISK	494	197	691	SPECIFICITY:	94%
	AT RISK	3	45	48	Fn RATE:	6%
	TOTAL	497	242	739	Fp RATE:	29%

 * The Screen for Grade 3 was not revalidated in 2008-2009 study; data for the Grade 3 Screen is drawn from the 2004-2006 revision.

WJ-Broad = the Woodcock-Johnson Broad Reading Cluster score. Risk was determined to be below the 20th percentile on the Broad Reading Cluster.