

**Technical Report # 27**

**Analysis of Reading Fluency and Comprehension Measures for Fourth  
Grade Students**

Julie Alonzo

Gerald Tindal

University of Oregon



behavioral research & teaching

Published by

Behavioral Research and Teaching  
University of Oregon • 175 Education  
5262 University of Oregon • Eugene, OR 97403-5262  
Phone: 541-346-3535 • Fax: 541-346-5689  
<http://brt.uoregon.edu>

Copyright © 2004. Behavioral Research and Teaching. All rights reserved. This publication, or parts thereof, may not be used or reproduced in any manner without written permission.

The University of Oregon is committed to the policy that all persons shall have equal access to its programs, facilities, and employment without regard to race, color, creed, religion, national origin, sex, age, marital status, disability, public assistance status, veteran status, or sexual orientation. This document is available in alternative formats upon request.

### *Introduction*

The No Child Left Behind Act of 2001 has increased the role of assessment in K-12 education. Designed to ensure that all students meet high academic standards, the law currently requires states receiving Title I funds to test all children annually in reading and math in grades 3 through 8 and report student performance disaggregated by poverty, race and ethnicity, disability, and limited English proficiency. By the 2005-06 school year, tests must be expanded to include at least one year between grades 10-12, and by 2007-08, states must also include science assessments at least once in grades 3-5, grades 6-9, and grades 10-12. The law requires states to set annual measurable objectives to track student progress towards reaching proficiency, with the ultimate goal that “all groups of students—including low-income students, students from major racial and ethnic groups, students with disabilities, and students with limited English proficiency—reach proficiency within 12 years” (U.S. Department of Education, 2002, p. 17).

With this goal in mind, school districts are scrambling to develop assessment systems that enable them to monitor student progress in a timely fashion rather than waiting for year-end statewide assessments. These district assessments serve multiple purposes: monitoring student progress, evaluating the effectiveness of particular programs and schools, and providing school personnel with valuable information about how well their students are doing. Developing easy-to-administer and score assessments at the district level offers schools a distinct advantage over depending on costly statewide assessments for progress monitoring. In the area of reading, three measures can provide essential information about students’ developing proficiency: oral reading fluency (ORF), vocabulary, and reading comprehension comprised of both selected response (SR) and constructed response (CR) items. Taken together, these three measures should give a good prediction of student performance on the large-scale reading assessment administered by

the state. In this technical report, data are presented on the technical adequacy of these measures as they are being developed, with an emphasis on predictive validity.

### *Methods*

#### *Setting and Subjects*

This report summarizes the spring 2003, fourth-grade reading achievement data from 29 different schools in an urban school district in the Pacific Northwest. The original data set contained 1,290 students, but some students were missing data in some but not all of the dependent variable measures, so the total sample size used for different analyses varies by measure.

#### *Design and Operational Procedures*

Dependent variables analyzed in this report include scores from the following measures: Oral Reading Fluency (ORF) ( $n = 1233$ ), a District Vocabulary Test ( $n = 1290$ ), a District Reading Comprehension Test ( $n = 1290$ ), and the previous year's statewide large-scale assessment in reading ( $n = 1097$ ). All fourth-grade students present in school on the days the tests were administered took all four assessments. Prior to analysis, schools in the district were coded into two regions, corresponding roughly with household income level. Independent blocking factors used in this report include income level, gender, ethnicity, and student status (Special Education [SPED] and English Language Learner [ELL] ).

#### *Measurement/Instrument Development*

##### *ORF*

The test of Oral Reading Fluency was administered individually to each student by trained assessors. Students read aloud for exactly one minute from one of two comparable passages deemed grade-level appropriate on the Flesch-Kincaid reading scale. At the end of one

minute, assessors marked the last word read then counted the total words read as well as any words read incorrectly to arrive at a final ORF score.

### *Vocabulary*

Fourth-grade students were administered a 25-word, multiple choice vocabulary test. Each item on the test consisted of one correct answer and two distracters. Students bubbled in their answers on the form itself, and all tests were machine scored.

### *Reading Comprehension*

Students were administered two reading comprehension tests. Each form of the reading comprehension test in which they read a passage and then answered multiple choice questions as well as two constructed response (CR) questions. The multiple choice can be considered as selected responses (SR) and were machine scored while CR questions were all scored by the test administrator using scoring guides provided by the district. The scorer was trained by two district administrators who also checked every fifth paper to ensure that scores were consistent with district expectations. When the scorer was unable to decide on an appropriate score, student responses were discussed with trainers before assigning a final score. Two different forms of the District Reading Comprehension Test were administered, varying in number of questions as well as genre of text passage read. This report includes suggestions for making the two forms more comparable in format as well as level of difficulty.

### *Oregon State Assessment in Reading*

In Oregon, students are administered the statewide exams in grades 3, 5, and 8. For this report, students' third-grade scores on the spring 2002 assessment in reading were used.

### *Data Preparation and Analysis*

The district ORF and reading comprehension test data were compared using analysis of variance (AOV) to check for comparability of forms and differential performance by different groups of students. For the District Vocabulary measure, Analysis of Variance (AOV) was used to test for differential performance by different groups of students. The percentage of students selecting each response was then calculated, along with the mean score on the measure for the students selecting each response, and the correlation between scores on the measure and response selected for each question. The Total Reading Scale Score on the statewide assessment was then correlated with all of the district measures and a multiple regression was used to ascertain optimal prediction from student performance on the four measures. Alpha was set at .05 for all analyses.

### *Results*

#### *ORF*

Table 1 presents descriptive statistics for the ORF. There was no statistically significant difference in student performance on the two different forms of the ORF  $F(1, 1226) = 1.45, p > .05$ , so ORF scores from both forms were combined for the rest of the analyses. Statistically significant differences were found in every comparison: (a) Females outperformed males; (b) Asians and Whites outperformed Hispanics; (c) general education students outperformed special education students; (d) non-ELL students outperformed ELL students; and (e) students from the higher income schools outperformed students from the lower income schools. It should be noted, however, that while these differences were statistically significant, the effect sizes were quite small, frequently accounting for only 1 – 2% of the variability of the groups sorted by gender, ethnicity, ELL and high-low income schools; the only practically significant variance was among

students designated as special education versus general education, which accounted for 10% of the variability in scores (see Table 2).

Table 1

*Descriptive Statistics for Grade 4 District ORF Test*

	Group	<i>n</i>	<i>M</i>	<i>SD</i>
Gender	Male	629	115.41	40.98
	Female	599	121.81	40.70
Ethnicity	White	816	122.35	40.35
	Hispanic	63	105.38	33.94
	African American	32	106.91	50.18
	Asian	58	129.91	34.37
	Native American	22	120.14	42.22
	Other	56	118.29	41.09
SPED		173	91.92	42.75
ELL		12	85.50	35.97
Income	Low	717	114.93	40.46
	High	511	123.58	41.14
Total		1228	118.53	40.95

Table 2

*Analysis of Variance Summary Table for Grade 4 District ORF Test*

Source	<i>df</i>	<i>F</i>	$\eta^2$	<i>p</i>
Gender	1	7.53**	.01	.01
Error	1226	(1668.23)		
Ethnicity	5	3.51**	.02	.00
Error	1041	(1608.28)		
SPED	1	119.12**	.10	.00
Error	1054	(1462.35)		
ELL	1	9.45**	.01	.00
Error	1054	(1613.16)		
Income	1	13.42**	.01	.00
Error	1223	(1660.29)		

*Note.* Values enclosed in parentheses represent mean square errors.

\* $p < .05$ , \*\* $p < .01$ .



*District Vocabulary Test*

Table 3 presents descriptive statistics for the District Vocabulary Test.

Table 3

*Descriptive Statistics for Grade 4 District Vocabulary Test*

	Group	<i>n</i>	<i>M</i>	<i>SD</i>
Gender	Male	643	79.86	26.46
	Female	607	82.53	24.60
Ethnicity	White	829	83.33	24.81
	Hispanic	64	72.06	25.78
	African American	32	85.38	16.65
	Asian	58	83.72	26.12
	Native American	23	67.30	37.79
	Other	57	84.49	23.41
SPED		172	72.42	26.15
ELL		12	56.00	20.61
Income	High	520	86.90	19.30
	Low	730	77.07	28.59
Total		1250	81.17	25.60

Statistically significant differences were found in every comparison except gender: (a) Asians, African Americans, and Whites outperformed Hispanics; (b) general education students outperformed special education students; (c) non-ELL students outperformed ELL students; and (d) students from the higher income schools outperformed students from the lower income

schools. While these differences were statistically significant, the effect sizes were quite small, accounting for only 1 – 4% of the variability between groups in all of the comparisons (see Table 4).

Table 4

*Analysis of Variance Summary Table for Grade 4 District Vocabulary Test*

Source	<i>df</i>	<i>F</i>	$\eta^2$	<i>p</i>
Gender	1	3.40	.00	.07
Error	1248	(0.07)		
Ethnicity	5	4.30**	.02	.00
Error	1057	(0.06)		
SPED	1	33.61**	.03	.00
Error	1070	(0.06)		
ELL	1	13.60**	.01	.00
Error	1070	(.06)		
Income	1	46.43**	.04	.00
Error	1274	(0.06)		

*Note.* Values enclosed in parentheses represent mean square errors.

\* $p < .05$ , \*\* $p < .01$ .

*District Reading Comprehension Test*

Table 5 presents descriptive statistics of student performance on the District Reading Comprehension Test. No significant difference was found between student performance on the selected response portion of the forms  $F(1, 1241) = 0.08, p > .05$ . A statistically significant difference was found between student performance on the constructed response portion of the two forms,  $F(1, 1241) = 191.19, p < .001$ .

Table 5

*Descriptive Statistics for Grade 4 District Reading Comprehension Test*

Form	<i>n</i>	SR <i>M</i>	SR <i>SD</i>	CR <i>M</i>	CR <i>SD</i>
A	591	73.18	19.20	37.92	14.66
B	652	72.89	16.80	56.77	30.05

Students performed at a significantly higher level on Form B (the *Basilisk* passage). This difference accounts for 13% of the variance in scores on the Constructed Response section of the District Reading Test. For this reason, Form A is separated out from Form B for analyses of student performance by group on the selected response section of the District Reading Comprehension Test (See Table 6).

Table 6

*Descriptive Statistics for Grade 4 District Reading Test: Selected Response Forms A and B*

	Group	<i>n</i>	<i>M</i>	<i>SD</i>
Gender	Male	638	72.28	18.03
	Female	605	73.82	17.89
Ethnicity	White	826	75.14	16.68
	Hispanic	63	65.14	18.74
	African American	31	68.33	20.23
	Asian	57	76.09	16.54
	Native American	23	66.07	19.31
	Other	57	73.34	17.29
SPED		169	65.03	18.94
ELL		12	47.94	18.10
Income	High	519	75.51	16.81
	Low	724	71.25	18.57
Total		1243	73.03	17.97

Significant differences were present in performance on the selected response section of the District Reading Comprehension Test between different groups of students on all of the blocking factors except gender: income level, ethnicity, and SPED or ELL designation. Students from high income schools outperformed students from low income schools, Asians and Whites outperformed Hispanics, and students not designated as special education or ELL outperformed their designated peers (See Table 7).

Table 7

*Analysis of Variance Summary Table for Grade 4 District Reading Test: Selected Response, Forms A and B*

Source	<i>df</i>	<i>F</i>	$\eta^2$	<i>p</i>
Gender	1	2.27	.00	.13
Error	1241	(.03)		
Ethnicity	5	6.04**	.03	.00
Error	1051	(.03)		
SPED	1	58.69**	.05	.00
Error	1064	(.03)		
ELL	1	28.84**	.03	.00
Error	1064	(.03)		
Income	1	17.20**	.01	.00
Error	1241	(.03)		

*Note.* Values enclosed in parentheses represent mean square errors.

\**p* < .05, \*\**p* < .01.

Table 8 presents descriptive statistics on student performance on the Constructed Response section of Form A of the District Reading Comprehension Test.

Table 8

*Descriptive Statistics for Grade 4 District Reading Test: Constructed Response, Form A*

	Group	<i>n</i>	<i>M</i>	<i>SD</i>
Gender	Male	286	36.93	15.43
	Female	305	38.85	13.85
Ethnicity	White	402	38.31	13.68
	Hispanic	31	34.68	12.79
	African American	13	37.50	21.65
	Asian	28	41.07	11.72
	Native American	6	33.33	23.27
	Other	20	41.25	16.27
SPED		81	31.64	18.66
ELL		7	37.50	7.22
Income	High	229	39.08	11.98
	Low	362	37.19	16.10
Total		591	37.92	14.66

A significant difference was found between special and general education students. This designation accounted for 4% of the overall variation in scores on the Constructed Response section of Form A of the District Reading Comprehension Test. No other significant differences were found between groups (see Table 9).

Table 9

*Analysis of Variance Summary Table for Grade 4 District Reading Test: Constructed Response, Form A*

Source	<i>df</i>	<i>F</i>	$\eta^2$	<i>p</i>
Gender	1	2.54	.00	.11
Error	589	(.02)		
Ethnicity	5	0.97	.01	.44
Error	494	(.02)		
SPED	1	22.64**	.04	.00
Error	504	(.02)		
ELL	1	0.02	.00	.88
Error	504	(.02)		
Income	1	2.35	.00	.13
Error	589	(.02)		

*Note.* Values enclosed in parentheses represent mean square errors.

\* $p < .05$ , \*\* $p < .01$

Table 10 presents descriptive statistics on student performance on the Constructed Response section of Form B of the District Reading Comprehension Test.

Table 10

*Descriptive Statistics for Grade 4 District Reading Test: Constructed Response, Form B*

	Group	<i>n</i>	<i>M</i>	<i>SD</i>
Gender	Male	352	53.80	29.84
	Female	300	60.25	29.97
Ethnicity	White	424	59.37	29.44
	Hispanic	32	49.61	33.22
	African American	18	46.53	24.93
	Asian	29	59.91	24.41
	Native American	17	41.91	24.98
	Other	37	60.14	33.96
SPED		88	44.32	31.48
ELL		5	35.00	29.84
Income	High	290	55.04	31.85
	Low	362	58.15	28.50
Total		652	56.77	30.05

Although the omnibus *F* test showed a statistically significant difference between student performance on the constructed response section of Form B of the Reading Comprehension Test based on ethnicity, post hoc analyses of results revealed no significant differences between the performance of different ethnic groups when unequal variances are accounted for. Levene's test of homogeneity of variances was significant for the constructed response section of the test, so



equal variances cannot be assumed; therefore, Tamhane's procedure was used for post-hoc comparison of performance. A significant difference was found between males and females as well as between students designated as SPED, those designated as ELL, and their general education or English first language peers, respectively. Gender and ELL designation each accounted for 1% of the overall variation in scores, while SPED designation accounted for 5% of the overall variation in scores on the Constructed Response section of Form B of the District Reading Comprehension Test (see Table 11).

Table 11

*Analysis of Variance Summary Table for Grade 4 District Reading Test: Constructed Response, Form B*

Source	<i>df</i>	<i>F</i>	$\eta^2$	<i>p</i>
Gender	1	7.54*	.01	.01
Error	650	(.09)		
Ethnicity	5	2.32*	.02	.04
Error	551	(.09)		
SPED	1	30.84**	.05	.00
Error	582	(.09)		
ELL	1	4.60*	.01	.03
Error	582	(.09)		
Income	1	1.48	.00	.22
Error	685	(.01)		

*Note.* Values enclosed in parentheses represent mean square errors.

\**p* < .05, \*\**p* < .01

*Correlation of the Four Measures*

Because student performance on the Constructed Response sections of Forms A and B of the District Reading Test differed significantly from each other, each form was considered separately for the remaining analyses.

*Correlations with Form A of CR District Reading Test*

Significant correlations existed between all of the measures analyzed in this study. The strongest correlation ( $r = .61$ ) was between the District ORF and the Statewide test in reading. Moderate positive correlations also existed between the District Vocabulary Test and the Statewide reading test ( $r = .44$ ) and between the SR section of the District Reading Test and the Statewide reading test ( $r = .51$ ). Table 12 presents the full results of these relationships.

Table 12

*Correlations Between the Grade 4 Measures, Form A of CR*

		District ORF	District Voc.	District SR Rdg	District CR Rdg	State Rdg
District ORF	Pearson Correlation	1	.363**	.444**	.332**	.605**
	Sig. (2-tailed)	.	.000	.000	.000	.000
	<i>n</i>	585	583	583	583	502
District Voc.	Pearson Correlation		1	.362**	.301**	.442**
	Sig. (2-tailed)		.	.000	.000	.000
	<i>n</i>		594	590	590	510
District SR Reading, Form A	Pearson Correlation			1	.195**	.513**
	Sig. (2-tailed)			.	.000	.000
	<i>n</i>			591	591	506
District CR Reading, Form A	Pearson Correlation				1	.344**
	Sig. (2-tailed)				.	.000
	<i>n</i>				591	506
State Reading	Pearson Correlation					1
	Sig. (2-tailed)					.
	<i>n</i>					511

\*\* . Correlation is significant at the .01 level (2-tailed).

*Correlations with Form B, of CR District Reading Test*

A significant correlation was found among all measures. The highest correlations were between the SR and CR sections of the District Reading Test ( $r = .63$ ), between the District ORF and the statewide test in reading ( $r = .61$ ), and between the SR section of the District Reading

Test and the Statewide test in reading ( $r = .59$ ). Table 13 presents full results for all correlational analyses.

Table 13

*Correlations Between the Grade 4 Measures, Form B of CR*

		District ORF	District Voc.	District SR Rdg	District CR Rdg	State Rdg
District ORF	Pearson Correlation	1	.326**	-.003	.457**	.606**
	Sig. (2-tailed)	.	.000	.939	.000	.000
	<i>n</i>	653	653	653	653	560
District Voc.	Pearson Correlation		1	.027	.402**	.264**
	Sig. (2-tailed)		.	.475	.000	.000
	<i>n</i>		687	687	687	584
District SR Reading, Form B	Pearson Correlation			1	-.047	-.008
	Sig. (2-tailed)			.	.223	.847
	<i>n</i>			687	687	584
District CR Reading, Form B	Pearson Correlation				1	.473**
	Sig. (2-tailed)				.	.000
	<i>n</i>				687	584
State Reading	Pearson Correlation					1
	Sig. (2-tailed)					.
	<i>n</i>					584

\*. Correlation is significant at the .05 level. \*\*. Correlation is significant at the .01 level (2-tailed).

*Regression Analysis of District Reading Assessments*

District ORF, Vocabulary, and Form A Reading Tests (both SR and CR) provide a statistically significant prediction of student performance on the previous spring's statewide

assessment in reading  $F(4, 495) = 120.88, p < .001$ . The district measures taken together (using Form A of the District Reading Test) account for 49% of the variability in state reading test performance, with ORF contributing the most to the explained variance (see Table 14).

Table 14

*Regression Summary for Grade 4 Statewide Reading Assessment, Form A*

Independent Variables	Unstandardized Coefficients		Standardized Coefficients	t	95% Confidence Interval for B	
	B	Std. Error	Beta		Lower Bound	Upper Bound
ORF	0.12	0.01	.400	10.72	0.10	0.14
District Vocabulary	11.80	2.16	.194	5.47	7.56	16.05
District Reading Test (Selected Response), Form A	16.72	2.56	.241	6.53	11.69	21.76
District Reading Test (Constructed Response), Form A	10.89	3.05	.122	3.58	4.91	16.88
Constant	176.22	2.12		83.07	172.05	180.39

District ORF and the District Reading Test (both SR and CR, Form B, also provide a statistically significant prediction of student performance on the previous spring's statewide assessment in reading  $F(4, 544) = 132.41, p < .001$ . Only vocabulary failed to contribute to the explained variance. These district measures taken together account for 49% of the variability in state reading test performance (see Table 15).

Table 15

*Regression Summary for Grade 4 Statewide Reading Assessment, Form B*

Independent Variables	Unstandardized Coefficients		Standardized Coefficients	t	95% Confidence Interval for B	
	B	Std. Error	Beta		Lower Bound	Upper Bound
ORF	0.13	0.01	0.40	11.12	0.11	0.15
District Vocabulary	-0.64	1.41	-0.02	-0.46	-3.41	2.13
District Reading Test (Selected Response), Form B	25.84	3.13	0.34	8.27	-3.87	3.52
District Reading Test (Constructed Response), Form B	4.51	1.64	0.11	2.74	19.70	31.98
Constant	182.34	1.91		95.30	178.58	186.09

*Discussion**ORF*

The ORF as it was administered in 2002-03 was strongly correlated with fourth-grade students' performance on the previous spring's statewide reading test ( $r = .61$ ) and moderately correlated with student performance on Form B of the District Reading Comprehension Test ( $r = .42 - .48$ ). A weaker correlation existed between the ORF and the CR section of Form A of the District Reading Test ( $r = .33$ ). Given its ease of administration and the fact that it does not require much time or training to score; this measure has continued to be a useful source of information for teachers monitoring student growth in reading, reflecting consistent outcomes with previous research for the past 20 years.

*District Vocabulary Test*

The District Vocabulary Test for Grade 4 was somewhat inconsistent between the two forms of the District Reading Test and would yield more useful information if it were made more challenging. Currently, it did not offer much insight into different levels of student reading skill because students were scoring over 80% correct on average.

*District Reading Comprehension Test*

The district administered two different forms of the Reading Comprehension test. One of the reading passages was non-fiction (Form A) and the other was fiction (Form B). Both forms had different numbers of questions and varied slightly in length and degree of difficulty on the Flesch-Kincaid reading scale. As discussed earlier, although both forms of the SR were similar in mean performance, a significant difference was found between student performance on the CR section of the two forms. Therefore, the CR section of the two forms were not comparable.

Recommendations to the district include the following. First, the constructed response forms need to be adjusted so they are more similar in difficulty. This can be accomplished by changing the questions (make them easier or more difficult accordingly) or changing the scoring rubric to reflect appropriate difficulty between the two forms. The district also can reduce the Selected Response section of each form to 15 questions and the Constructed Response section of each form to 2 questions. Following are suggestions for removal of items to shorten the forms on the Selected Response and make the two forms of the Constructed Response statistically insignificant. Table 16 presents recommendations for item removal based on an analysis of how the different items were functioning. To make the Constructed Response section of the forms more comparable, the district needs to re-write Question #20 to make it more challenging or re-write the scoring rubric to allow for more differentiation between scores, as a two point scale

with only two questions leaves little room for variation in performance. Furthermore, two questions should be removed from Form B.

Table 16

*Items for Removal from Grade 4 Reading Test and How Removal Would Affect Scores*

Form	SR Item #s for Removal	New Mean SR Score	SR Score Before Removal	CR Item #s for Removal	New Mean CR Score	CR Score Before Removal
A	3, 4, 14	71	73	none	37	38
B	3, 13, 16, 19, 20, 21, 22	73	67	25, 26	26	58

These recommendations in Table 16 are based on student performance; in Table 17 a rationale is provided for removing these items. In this latter table, an item is considered *redundant* if students performed equally well on that item as they did on another item on the same form. The % given in parentheses refers to the percentage of fourth-grade students who got that particular item correct. The *Action Needed to Save Item for Question Bank* column indicates which questions the district can retain with confidence to use in future tests, and—where appropriate—what the district needs to do in order to make the item more usable. See Appendix A for a complete table of Item Analysis for the Selected Response section of the District Reading Test.



Table 17

*Rationale for Items Suggested for Removal from Grade 4 District Reading Test*

Form	Item	Rationale for Removal	Action Needed to Save Item for Question Bank
A	3	Too easy (90%)	Re-write question to make it more challenging
A	4	Too easy (93%)	Re-write question to make it more challenging
A	14	Redundant with Item 16	OK to use as is in place of Item 16
B	3	Too easy (92%)	Re-write question to make it more challenging
B	13	Redundant with Item 9 and 22	OK to use as is in place of Item 9 or 22
B	16	Redundant with Item 1	OK to use as is in place of Item 1
B	19	Redundant with Item 12	OK to use as is in place of Item 12
B	20	Redundant with Item 14	OK to use as is in place of Item 14
B	21	Too hard (42%)	Re-write question to make it less challenging
B	22	Redundant with Item 9 and 13	OK to use as is in place of Item 9 or 13
B	25	Redundant with Item 23 and 24	OK to use as is in place of Item 23 or 24
B	26	Redundant with Item 23 and 24	OK to use as is in place of Item 23 or 24

The district's current reading assessment kit can offer insights into strengths of particular programs, schools, and teachers and provide school personnel with information that can help them measure their progress towards promoting reading proficiency for all students. It will continue to be revised, and the revisions will be analyzed using Item Response Theory (IRT) in subsequent years as the district works to improve the reliability and validity of the instruments for the various ways they are used. Additional technical reports will be written to follow up on these analyses and document the changes being made to the reading assessment kit.

*References*

U.S. Department of Education. (2002). *No child left behind: a desktop reference*. U.S. Department of Education, Office of Elementary and Secondary Education. Washington, DC: Author.

## Appendix A

*Item Analysis for District Reading Test, Selected Response: Forms A and B*

Item	Form	% of students who got item correct	% of students selecting Option A	% of students selecting Option B	% of students selecting Option C	% of students selecting Option D	Mean score of students selecting Option A	Mean score of students selecting Option B	Mean score of students selecting Option C	Mean score of students selecting Option D
1	A	0.72	0.08	0.08	0.53	0.31	0.32	0.51	0.72	0.89
2	A	0.81	0.09	0.03	0.34	0.53	0.72	0.56	0.79	0.69
3	A	0.90	0.58	0.38	0.03	0.01	0.70	0.77	0.56	0.60
4	A	0.93	0.41	0.01	0.57	0.02	0.76	0.63	0.70	0.67
5	A	0.61	0.61	0.15	0.17	0.06	0.74	0.68	0.70	0.68
6	A	0.83	0.10	0.82	0.05	0.02	0.67	0.73	0.69	0.80
7	A	0.73	0.06	0.06	0.14	0.72	0.64	0.67	0.73	0.73
8	A	0.88	0.50	0.02	0.42	0.06	0.69	0.68	0.77	0.71
9	A	0.82	0.08	0.07	0.02	0.81	0.73	0.74	0.68	0.72
10	A	0.70	0.06	0.30	0.60	0.03	0.71	0.70	0.73	0.73
11	A	0.57	0.38	0.31	0.16	0.15	0.70	0.77	0.69	0.72
12	A	0.78	0.37	0.48	0.03	0.10	0.78	0.69	0.65	0.69
13	A	0.79	0.04	0.12	0.40	0.43	0.65	0.67	0.77	0.70
14	A	0.76	0.37	0.07	0.49	0.07	0.76	0.65	0.70	0.74
15	A	0.55	0.06	0.58	0.32	0.04	0.67	0.74	0.71	0.69
16	A	0.70	0.45	0.12	0.04	0.38	0.70	0.65	0.64	0.79
17	A	0.66	0.37	0.02	0.13	0.47	0.70	0.57	0.70	0.75
18	A	0.54	0.09	0.48	0.22	0.20	0.70	0.71	0.75	0.74

1	B	0.68	0.13	0.08	0.52	0.26	0.68	0.69	0.67	0.67
2	B	0.83	0.09	0.03	0.37	0.50	0.70	0.73	0.66	0.67
3	B	0.92	0.55	0.41	0.03	0.01	0.68	0.67	0.69	0.59
4	B	0.93	0.43	0.02	0.54	0.01	0.67	0.64	0.67	0.81
5	B	0.65	0.65	0.15	0.13	0.07	0.67	0.69	0.66	0.67
6	B	0.85	0.08	0.84	0.05	0.02	0.68	0.67	0.62	0.75
7	B	0.73	0.05	0.08	0.14	0.73	0.69	0.69	0.68	0.67
8	B	0.90	0.50	0.02	0.43	0.05	0.67	0.65	0.67	0.71
9	B	0.81	0.09	0.08	0.03	0.80	0.67	0.64	0.68	0.68
10	B	0.64	0.07	0.27	0.60	0.05	0.65	0.68	0.67	0.69
11	B	0.60	0.40	0.34	0.12	0.14	0.67	0.67	0.70	0.65
12	B	0.76	0.35	0.47	0.06	0.12	0.67	0.67	0.66	0.69
13	B	0.81	0.04	0.12	0.42	0.42	0.64	0.69	0.67	0.67
14	B	0.82	0.39	0.04	0.51	0.05	0.68	0.69	0.67	0.68
15	B	0.58	0.07	0.60	0.30	0.04	0.69	0.66	0.69	0.65
16	B	0.68	0.40	0.16	0.04	0.40	0.68	0.67	0.64	0.67
17	B	0.67	0.33	0.02	0.12	0.52	0.67	0.72	0.72	0.66
18	B	0.55	0.08	0.45	0.24	0.22	0.63	0.68	0.67	0.67
19	B	0.76	0.01	0.41	0.05	0.07	0.68	0.67	0.66	0.70
20	B	0.82	0.01	0.06	0.02	0.44	0.73	0.70	0.70	0.67
21	B	0.42	0.23	0.04	0.23	0.05	0.68	0.70	0.66	0.65
22	B	0.81	0.05	0.44	0.03	0.02	0.69	0.68	0.60	0.64

---