



# **Examining Differential Item Functioning in Reading Assessments for Students with Disabilities**





**Partnership for Accessible Reading Assessment**

University of Minnesota  
350 Elliott Hall  
75 E. River Road  
Minneapolis, MN 55455

<http://www.readingassessment.info>  
[readingassess@umn.edu](mailto:readingassess@umn.edu)

This work is supported, in part, by the U.S. Department of Education, Institute of Education Sciences, National Center for



Special Education Research—Grant No. H324F040002. Opinions expressed do not necessarily reflect those of the U.S. Department of Education or offices within it.



UNIVERSITY  
OF MINNESOTA  
**Curriculum and  
Instruction**



**UC DAVIS**



The University of Minnesota 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.

# Examining Differential Item Functioning in Reading Assessments for Students with Disabilities

**Jamal Abedi**

**National Center for Research on Evaluation, Standards, & Student Testing  
University of California, Davis**

**Seth Leon & Jenny C. Kao**

**National Center for Research on Evaluation, Standards, & Student Testing  
University of California, Los Angeles**

January 2007

All rights reserved. Any or all portions of this document may be reproduced and distributed without prior permission, provided the source is cited as:

Abedi, J., Leon, S., & Kao, J. (2007). *Examining differential item functioning in reading assessments for students with disabilities*. Minneapolis, MN: University of Minnesota, Partnership for Accessible Reading Assessment.



**Partnership for Accessible Reading Assessment**

University of Minnesota  
350 Elliott Hall  
75 E. River Road  
Minneapolis, MN 55455

<http://www.readingassessment.info>  
[readingassess@umn.edu](mailto:readingassess@umn.edu)

This work is supported, in part, by the U.S. Department of Education, Institute of Education Sciences, National Center for Special Education Research—Grant No. H324F040002. Opinions expressed do not necessarily reflect those of the U.S. Department of Education or offices within it.



The University of Minnesota 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.

## Abstract

This study examines group differences between students with disabilities and students without disabilities using DIF analyses in a high-stakes reading assessment. Results indicated that for grade 9, many items exhibited DIF. Items that exhibited DIF were more likely to be located in the second half of the assessment subscales. After accounting for reading ability using a proxy score from items on the first half of the subscales, students with disabilities consistently under-performed on items located in the second half relative to the items located in the first half, compared to students without disabilities. These results were seen in grade 9 for data from two different states, but these results were not seen for grade 3. This study has several limitations to the data. There was no access to information about the testing accommodations that students with disabilities might have received, and no information about the type of disabilities. Results of this study can shed light on potential factors affecting the accessibility of reading assessments for students with disabilities, in an ultimate effort to provide assessment tools that are conceptually and psychometrically sound for all students. A companion report is available examining differential distractor functioning for students with disabilities.



## Table of Contents

Introduction.....	1
Research Questions.....	2
Methodology.....	3
Data Source.....	3
Procedure & Statistical Design.....	4
Results.....	4
State X Grade 9.....	5
State Y Grade 9.....	9
State X Grade 3.....	16
State Y Grade 3.....	19
Discussion.....	22
References.....	25
Acknowledgements.....	27
Appendix. Detailed DIF Results.....	29





## Introduction

More than 6 million students with disabilities—approximately 13 percent of all students—attended United States public schools during the 2003–2004 school year (U.S. Government Accounting Office, 2005). Accountability standards have been raised since the reauthorization of the Individuals with Disabilities Education Act (IDEA) and the authorization of the No Child Left Behind Act of 2001 (NCLB, 2002), that require states to include students with disabilities in annual assessments. In a review of state practices, Klein, Wiley, and Thurlow (2006) found that 44 states reported participation and performance for students with disabilities on all of their NCLB assessments during the 2003–2004 school year. According to data collected during the 2003–2004 school year, of the 48 reporting states and the District of Columbia, 41 states reported that at least 95 percent of students with disabilities participated in the statewide reading assessment (U.S. Government Accounting Office, 2005). Furthermore, most students with disabilities participated in regular reading assessments, while relatively few participated in alternate assessments.

Nearly 84% of middle school students with Individualized Education Programs (IEPs) participated in general reading assessments, as reported by states in the 2002–2003 Annual Performance Reports (Thurlow, Moen, & Wiley, 2005). Given the high rate of participation by students with disabilities in regular state and national assessments, as well as the implications of assessment outcomes for accountability, it is imperative that we ensure these assessments are as accessible to students with disabilities as possible. In other words, they must be as fair and accurate as possible. Students with disabilities may perform less well than their peers without disabilities for a variety of reasons, including their specific disability, lack of appropriate testing accommodations, or lack of opportunity to learn. However, they may also perform less well because of factors directly related to the tests. For instance, there could be issues related to the item quality or test item format. It is necessary to reduce irrelevant and extraneous sources not related to the construct being measured.

Test bias can occur when performance on a test requires sources of knowledge different from those intended to be measured, causing test scores to be less valid for a particular group (Penfield & Lam, 2000). Test bias is often examined at the item level, with differential item functioning (DIF) analyses being part of the framework for probing item bias. If a certain group (i.e., racial/ethnic group or gender) performs lower on average on a specific item, then one could say that the item is biased against that particular group. DIF analyses compare the performance of two groups of the same level of ability in order to disentangle the effects of unfairness and ability level. Matching ability level is essential, since different groups may have different ability levels, where case differences in performance are to be expected (Clauser & Mazor, 1998). Consistent differences between two groups of the same ability level would suggest that DIF is present. However, results of DIF analyses can only suggest that DIF is present, and not that the items are biased. To consider an item as biased also requires determining the non-target constructs that lead

to the between-group differences in performance (Penfield & Lam, 2000). Thus, DIF is a necessary but not sufficient condition for item bias (Clauser & Mazor, 1998).

DIF analysis is often used to examine group differences between specific racial or ethnic groups or between males and females. For example, Hauser and Kingsbury (2004) explored differential functioning across student groups formed based on ethnicity and based on gender on items from the Idaho Standards Achievement Test. Zenisky, Hambleton, and Robin (2004) explored gender DIF in a large-scale science assessment. Other research has examined incidences of DIF for limited English proficient students (Snetzler & Qualls, 2000). DIF analyses have also been conducted for students with disabilities. Specifically, DIF analyses have been used to examine effects of accommodations that are provided to students with disabilities during testing (Bolt, 2004; Cohen, Gregg, & Deng, 2005; Koretz & Hamilton, 1999).

This study aims to examine potential factors that may affect the accessibility of reading assessments for students with disabilities. Haladyna and Downing (2004) identified potential sources of systematic errors associated with construct-irrelevant variance, that included factors relating to test development: (1) item quality; (2) test item format; and (3) differential item functioning. We were specifically interested in employing DIF analyses to examine any potential between-groups differences in a high-stakes reading assessment. Our study differs from previous research using DIF analyses for students with disabilities in that our study seeks to investigate specific factors related to the test rather than to the accommodation.

There are several statistical procedures that can be used to identify differentially functioning test items, including the Mantel-Haentzel statistic, logistic regression, SIBTEST, the Standardization procedure, and various Item-Response-Theory-based approaches (Clauser & Mazor, 1998). Our study uses a logistic regression approach as outlined by Zumbo (1999) because it is easier to employ and is more suitable for answering our research questions.

## Research Questions

The following research questions guided the analyses and reporting of this study:

1. Do items on standardized Reading Comprehension (RC) and Word Analysis (WA) subscales exhibit Differential Item Functioning (DIF) for students with disabilities?
2. Are more items that exhibit DIF for students with disabilities located in the second half of RC and WA subscales rather than in the first half?

3. Do students with disabilities consistently under-perform on items located in the second half relative to items located in the first half, as compared to students without disabilities?
4. Do the results of DIF vary by grade (grade 3 and grade 9)?

## Methodology

### Data Source

Data from two states provided the impetus for answering the research questions. We will refer to them as State X and State Y to ensure anonymity.

State X is a small state with an average number of students with disabilities. Data were obtained for the 1997–1998 academic year and included item-level information on students’ responses on the Stanford Achievement Test, Ninth Edition (Stanford 9). Students with valid scores were included in our analyses. Students with LEP (limited English proficient) classifications (including LEP students with disabilities) were excluded from the analyses to reduce the possible confounding of language proficiency issues. Of the 6,611 third-grade students included in the present analyses, 448 (6.8%) were considered to be students with disabilities. Of the 5,287 ninth-grade students, 522 (9.9%) were considered to be students with disabilities.

State Y is a large state with an average number of students with disabilities. Data were obtained for the 1997–1998 academic year and included item-level information on students’ responses in the Stanford 9. Students with valid scores were included in our analyses. Students with LEP classifications (including LEP students with disabilities) were excluded from the analyses to reduce the possible confounding of language proficiency issues. Of the 278,287 third-grade students included in the present analyses, 21,239 (7.6%) were considered to be students with disabilities. Of the 244,446 ninth-grade students, 17,321 (7.1%) were considered to be students with disabilities.

Published by Harcourt Brace Educational Measurement in 1996, the Stanford 9 is a standardized, norm-referenced test in several subject areas, including reading. According to the Harcourt Assessment website, the Stanford 9 uses an “easy-hard-easy format” in which “difficult questions are surrounded by easy questions to encourage students to complete the test.” The reading portion of the test is characterized by three different types of reading selections—recreational, textual, and functional—and items that assess initial understanding, interpretation, critical analysis, and reading strategy (HarcourtAssessment.com).

The present study examines two subscales of the Stanford 9, Reading Comprehension (RC) and Word Analysis (WA) (more commonly known as “phonics” or “decoding”), from the two states. Public school students in grades 3 and 9 were analyzed to present data over a wider age range.

## Procedure & Statistical Design

To determine whether items exhibit DIF for students with disabilities, a multi-step logistic regression procedure was employed. The outcome variable in each model was the dichotomous response to the item which was coded as correct or incorrect. A total score on the applicable subscale (RC or WA) was computed as a proxy for ability on the construct. In step 1, the ability proxy was entered into the model and a measure of the explained variance (Nagelkerke R-square) was obtained. In step 2 the disability status grouping variable and an interaction between disability status and the ability proxy were entered into the model. Again the R-square estimate was obtained. The change in R-square between step 1 and step 2 was calculated and tested for significance. Items were identified for closer inspection as differentially functioning if the R-square change was at least 0.003 and was significant at  $p < 0.01$ .

A similar approach was used to determine whether item order influences DIF for students with disabilities. Rather than using the total score as a proxy for ability only the score on items from the first half of the assessment was used as an ability proxy (first 27 out of 54 items for RC; first 15 out of 30 items for WA). Items that exhibited DIF were examined more closely looking at the odds ratios of the variables in the final model. If systemic differences in the DIF findings arose between the two approaches they could then be compared. For example, if items showed larger DIF effects on the items from the latter portion of the assessment when the second proxy was used, and if the odds ratios on those items were in a consistent direction, then it would be apparent that item order was influencing DIF.

Logistic regression was selected as the statistical procedure due to its ability to detect both uniform and non-uniform DIF and its ease of availability on the SPSS platform. A main effect of the ability proxy would be an indication of uniform DIF whereas a significant effect with the addition of the interaction between disability status and the ability proxy would suggest non-uniform DIF.

## Results

The analyses examine the following research questions:

1. Do items on standardized Reading Comprehension (RC) and Word Analysis (WA) subscales exhibit Differential Item Functioning (DIF) for students with disabilities?
2. Are more items that exhibit DIF for students with disabilities located in the second half of RC and WA subscales rather than in the first half?

3. Do students with disabilities consistently under-perform on items located in the second half relative to items located in the first half, as compared to students without disabilities?
4. Do the results of DIF vary by grade (grade 3 and grade 9)?

The results are described by state and grade, and then by subscale. Detailed results of the DIF findings are available in the Appendix.

### State X Grade 9

**Reading Comprehension.** Table 1 presents DIF results from State X in grade 9 for the 54-item Reading Comprehension subscale. The total score on the 54 items served as an ability proxy in this model. Items were identified as differentially functioning when the R-square change between steps 1 and 2 was at least 0.003 and was significant at  $p < 0.01$ . There were 17 items that showed DIF, 13 of which were located in the second half of the assessment (Items 28 through 54). This suggests that item order might be influencing DIF.

The second model used a similar method with the exception that the ability proxy was calculated only from the first 27 items. Using this method there were 23 items that showed DIF, 17 of which came from the second half of the assessment. The effect sizes using the first half ability proxy were larger, especially for the items from the second half of the assessment.

**Table 1. State X Grade 9 Item-level Reading Comprehension**

Ability Proxy	Total Number of Items	Number of Items Showing DIF		
		Items 1–27	Items 28–54	All Items
Model 1	54	4	13	17
Model 2	54	6	17	23

*Note:* In Model 1, the total score was used as an ability proxy. In Model 2, the score on the first 27 items was used as an ability proxy.

Items that were found to exhibit DIF from Model 2 in Table 1 were examined more closely in Table 2 to determine whether item order might be systematically influencing DIF. Logistic regression models were re-run for each of the 17 DIF items from the second half of the test. Each of the three variables was entered in a separate step to determine each partial R-square addition. Odds ratios are presented for the full model. In 15 of the 17 items the main effect of the disability status grouping variable was significant and for all 15 of those items the odds ratio for the disability status grouping variable was less than 1.0. This strongly demonstrates that students with disabilities under-performed on each of those items relative to students without disabilities when controlling for per-

formance on the first half of the assessment. Similarly, 14 of these 17 items had a significant interaction between the disability status grouping variable and the first half ability proxy and the odds ratio for each significant finding was less than 1.0. A significant interaction term with an odds ratio less than 1.0 indicates that a student with disabilities who scored well on the first 27 items would not score as well on the second half of the test as a student without disabilities who had scored similarly on the first 27 items.

**Table 2. State X Grade 9 Item-level Reading Comprehension Logistic Regression Results for Items Showing DIF with Ability Proxy Based On First 27 Items Score**

Item No.	R-square results at each step in the sequential logistic regression			Odds Ratios–Final Model		
	Step 1 Ability Proxy	Step 2 Ability Proxy and disability status  (Uniform)	Step 3 Ability Proxy, disability status and interaction  (Non-Uniform)	Ability Proxy	Disability Status	Interaction
29	0.176**	0.176	0.179**	2.41**	0.66**	0.68**
32	0.171**	0.177**	0.179**	2.25**	0.45**	0.73**
33	0.195**	0.196	0.201**	2.59**	0.60**	0.60**
35	0.034**	0.034	0.037**	1.46**	0.74*	0.73**
36	0.065**	0.073**	0.073	1.51**	0.51**	0.97
37	0.222**	0.222	0.225**	2.76**	0.68**	0.67**
40	0.226**	0.226	0.229**	2.80**	0.71**	0.69**
41	0.114**	0.114	0.121**	2.12**	0.72*	0.58**
42	0.098**	0.098	0.106**	2.03**	0.74*	0.55**
43	0.250**	0.252**	0.254*	2.86**	0.58**	0.77*
44	0.161**	0.165**	0.168**	2.23**	0.46**	0.66**
45	0.140**	0.143**	0.144	2.06**	0.57**	0.83
48	0.144**	0.154**	0.155	1.91**	0.55**	1.12
49	0.209**	0.211**	0.218**	2.82**	0.46**	0.50**
51	0.101**	0.101	0.104*	1.98**	0.80	0.67**
52	0.127**	0.129**	0.132**	2.33**	1.05	0.64**
54	0.230**	0.237**	0.240**	2.67**	0.40**	0.67**

Note: \* denotes significance at  $p < .05$ . \*\* denotes significance at  $p < .01$

Figures 1 and 2 of the expected probability of a correct response for Items 36 and 49, respectively, serve as examples to illustrate these respective effects.

**Figure 1. Expected Probability of a Correct Response for Item 36 in State X Grade 9 Reading Comprehension**

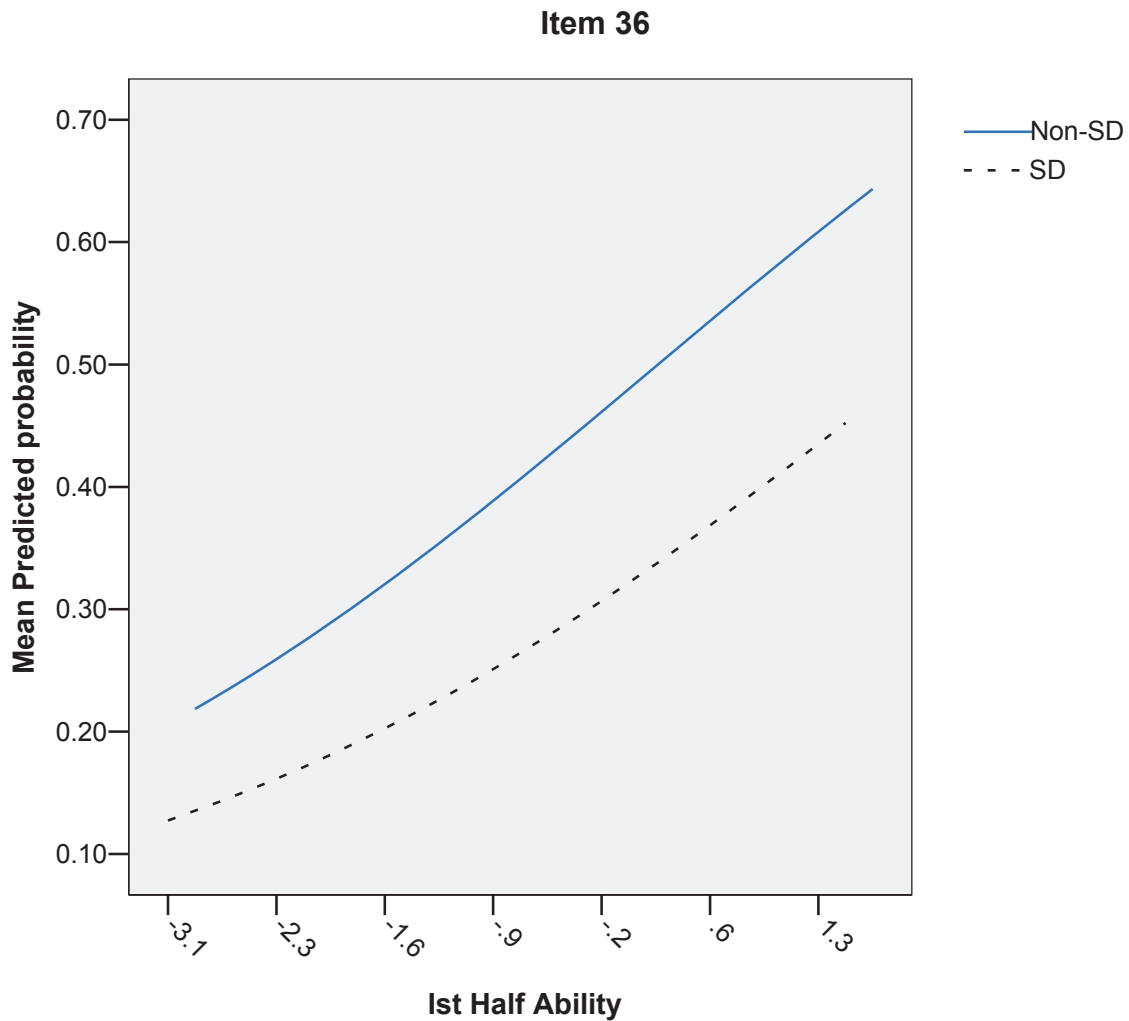
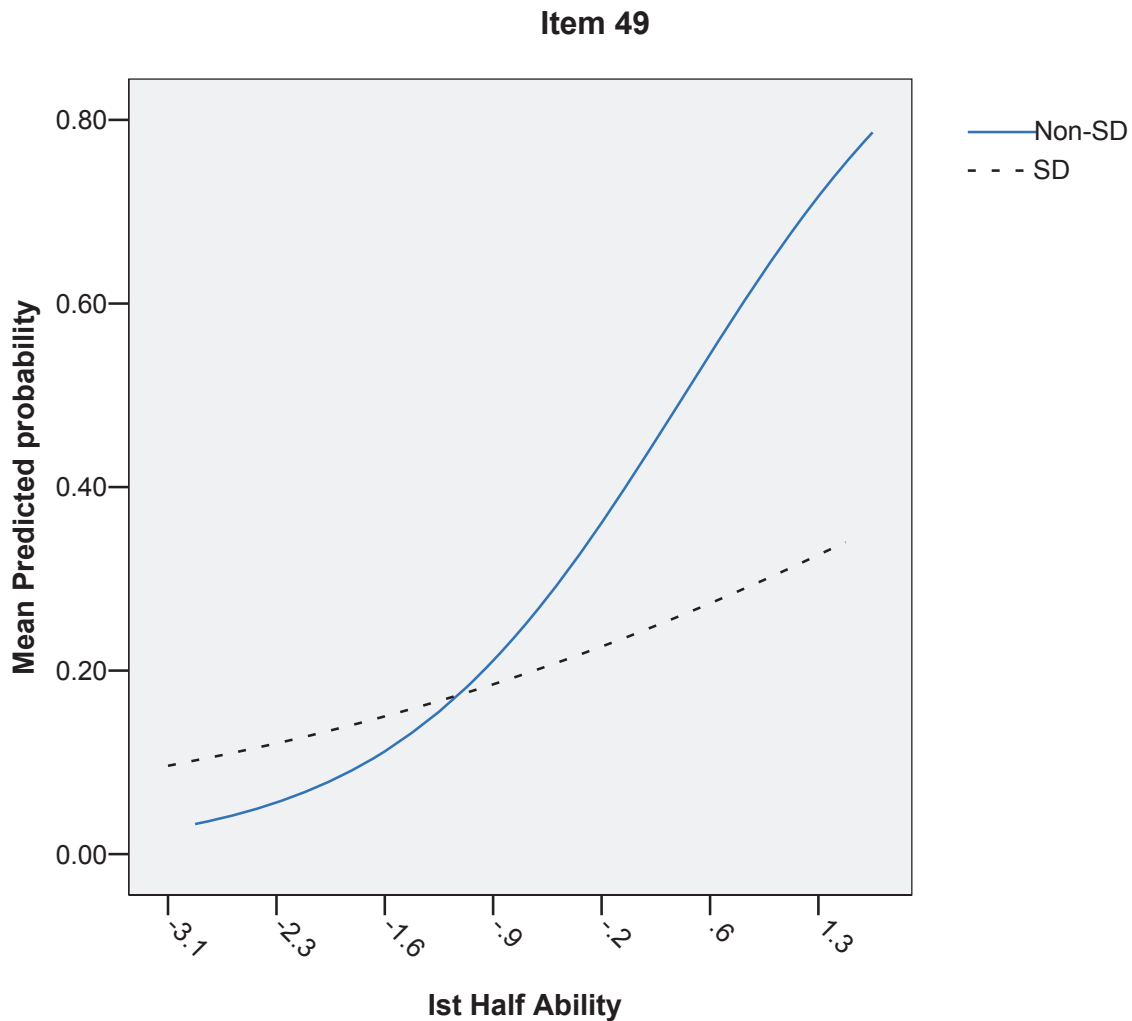


Figure 1 represents the relationship for a strong main effect on the disability status grouping variable. The odds ratio for the main effect of the disability status grouping variable was 0.51. Students with disabilities who scored similarly as students without disabilities on the first half of the assessment were less likely to answer Item 36 correctly.

Figure 2 represents the relationship for an interaction between the disability status grouping variable and the ability proxy based on the score from the first half. The odds ratio for the interaction term on item 49 was 0.5. The performance gap between students

with disabilities and students without disabilities becomes very large for students who performed well on the first half of the test and there is little gap for students who were one standard deviation or more below the mean on the first half of the assessment.

**Figure 2. Expected Probability of a Correct Response for Item 49 in State X Grade 9 Reading Comprehension**



**Word Analysis.** Table 3 presents DIF results from State X in grade 9 for the 30-item Word Analysis subscale. The total score on the 30 items served as an ability proxy in this model. Items were identified as DIF when the R-square change between steps 1 and 2 was at least 0.003 and was significant at  $p < 0.01$ . There were 12 items that showed DIF, 8 of which were located in the second half of the assessment (Items 16 through 30). Similar to the results for RC, this suggests that item order might be influencing DIF.



The second model used a similar method with the exception that the ability proxy was calculated only from the first 15 items. Using this method there were 19 items that showed DIF, 13 of which came from the second half of the assessment. The effect sizes using the first half ability proxy were larger, especially for the items from the second half of the assessment.

**Table 3. State X Grade 9 Item-level Word Analysis**

Ability Proxy	Total Number of Items	Number of Items Showing DIF		
		Items 1–15	Items 16–30	All items
Model 1	30	4	8	12
Model 2	30	6	13	19

*Note:* In Model 1, the total score was used as an ability proxy. In Model 2, the score on the first 15 items was used as an ability proxy.

Items that were found to exhibit DIF from Model 2 in Table 3 were examined more closely in Table 4 to determine if item order might be systematically influencing DIF. Logistic regression models were re-run for each of the 13 DIF items from the second half of the test. Each of the three variables was entered in a separate step to determine each partial R-square addition. Odds ratios are presented for the full model. In 12 of the 13 items the main effect of the disability status grouping variable was significant and for all 12 of those items the odds ratio for the disability status grouping variable was less than 1.0. Again this strongly demonstrates that students with disabilities under-performed on each of those items relative to students without disabilities when controlling for performance on the first half of the assessment. Additionally, 5 of these 13 items had a significant interaction between the disability status grouping variable and the first half ability proxy and the odds ratio for each significant finding was less than 1.0. All five significant interaction effects occurred on items located near the end of the test.

**Table 4. State X Grade 9 Item-level Word Analysis Logistic Regression Results for Items Showing DIF with Ability Proxy Based On First 15 Items Score**

Item	R-square Results at Each Step in the Sequential Logistic Regression			Odds Ratios–Final Model		
	Step 1 Ability Proxy	Step 2 Ability Proxy and Disability Status (Uniform)	Step 3 Ability Proxy, Disability Status and Interaction (Non-Uniform)	Ability Proxy	Disability Status	Interaction
16	0.145**	0.151**	0.151	2.06**	0.56**	0.89
17	0.071**	0.075**	0.075	1.60**	0.71**	1.08
18	0.129**	0.147**	0.147	1.99**	0.39**	0.84
19	0.105**	0.125**	0.125	1.83**	0.43**	1.04
20	0.029**	0.043**	0.043	1.26**	0.53**	1.15
21	0.196**	0.207**	0.208	2.53**	0.43**	0.84
22	0.49**	0.53**	0.53	1.46**	0.65**	0.95
23	0.153**	0.169**	0.169	2.14**	0.39**	0.83
24	0.194**	0.197*	0.199**	2.49**	0.59**	0.72**
25	0.201**	0.202	0.209**	2.78**	0.82	0.52**
27	0.201**	0.211**	0.215**	2.52**	0.38**	0.65**
28	0.180**	0.188**	0.190*	2.42**	0.46**	0.75**
30	0.266**	0.288**	0.290**	3.30**	0.28**	0.71**

Note: \* denotes significance at  $p < .05$ . \*\* denotes significance at  $p < .01$

Figures 3 and 4 of the expected probability of a correct response for Items 18 and 30, respectively serve as examples to illustrate these respective effects.

**Figure 3. Expected Probability of a Correct Response for Item 18 in State X Grade 9 Word Analysis**

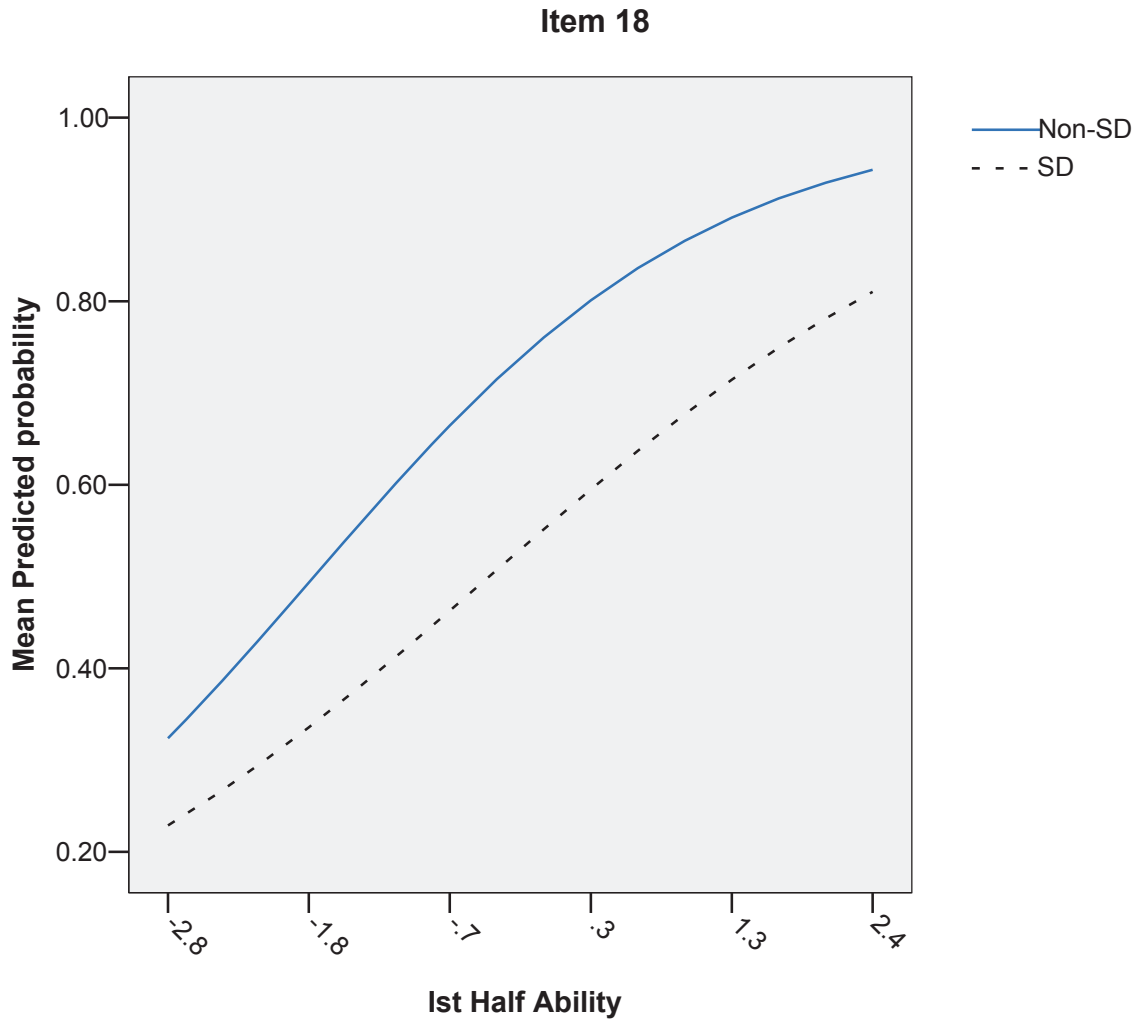
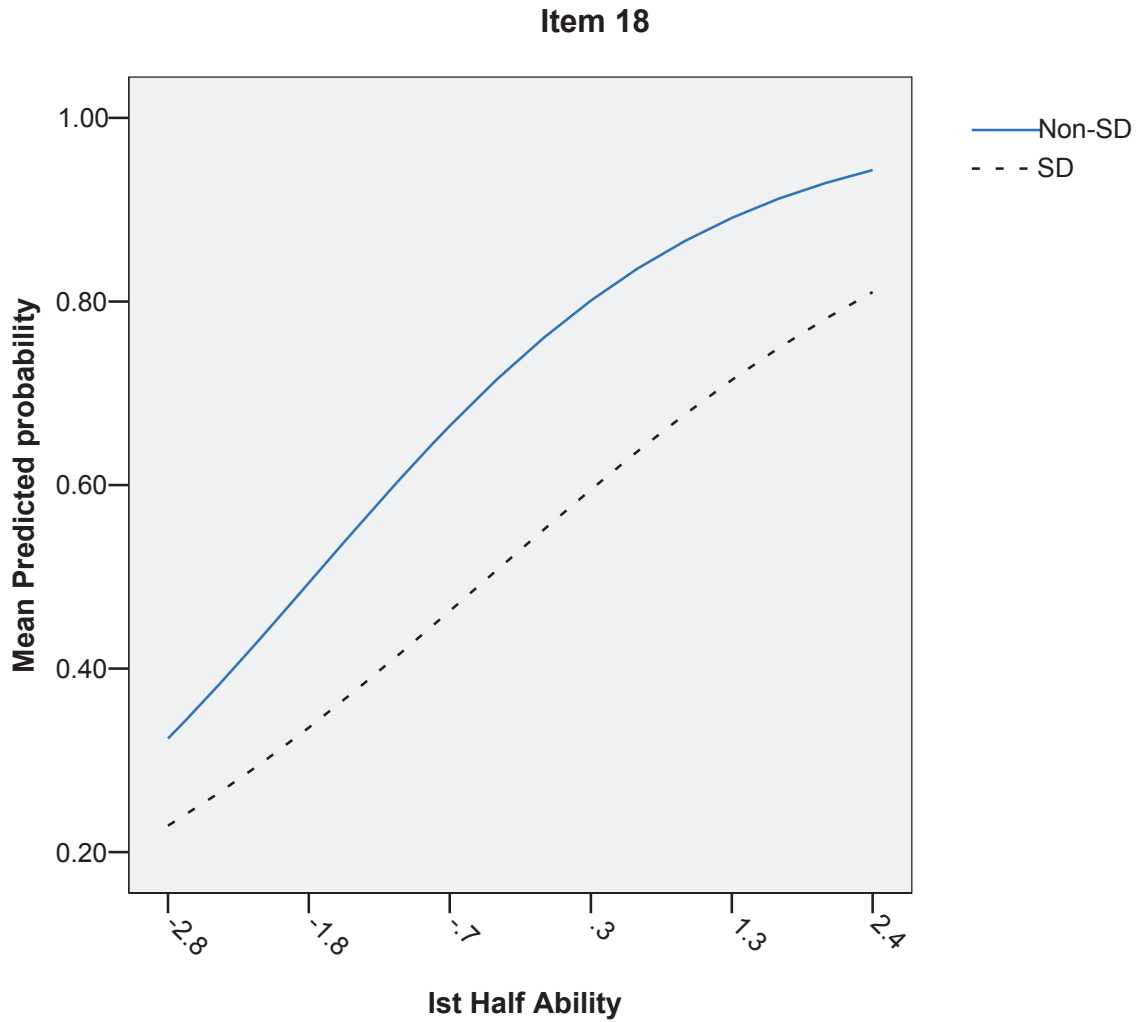


Figure 3 represents the relationship for a strong main effect on the disability status grouping variable. The odds ratio for the main effect of the disability status grouping variable was 0.39. Students with disabilities who scored similarly to students without disabilities on the first half of the assessment were less likely to answer Item 18 correctly.

Figure 4 represents the relationship for an interaction between the disability status grouping variable and the ability proxy based on the score from the first half, along with a strong main disability effect. The odds ratio for the interaction term on Item 30 was 0.71. The odds ratio for the main disability effect was 0.28. Students with disabilities with similar performance to students without disabilities on the first 15 items are always pre-

dicted to score below non-disabled students. The gap between students with disabilities and students without disabilities in expected performance increases as performance on the first half of the test increases.

**Figure 4. Expected Probability of a Correct Response for Item 30 in State X Grade 9 Word Analysis**



## State Y Grade 9

**Reading Comprehension.** Table 5 presents DIF results from State Y in grade 9 for the 54-item Reading Comprehension subscale. Items were identified as DIF when the R-square change between steps 1 and 2 was at least 0.003 and was significant at  $p < 0.01$ . When using total score on the 54 items as an ability proxy there were no items that showed DIF although most items were significant at  $p < 0.01$ .

In the second model, in which the ability proxy was calculated only from the first 27 items, 13 items showed DIF, 11 of which were located in the second half of the assessment. The effect sizes using the ability proxy based on the score from the first half were larger, especially for the items from the second half of the assessment.

**Table 5. State Y Grade 9 Item-level Reading Comprehension**

Ability Proxy	Total Number of Items	Number of Items Showing DIF		
		Items 1–27	Items 28–54	All Items
Model 1	54	0	0	0
Model 2	54	2	11	13

*Note:* In Model 1, the total score was used as an ability proxy. In Model 2, the score on the first 27 items was used as an ability proxy.

Items that were found to exhibit DIF from Model 2 in Table 5 were examined more closely in Table 6 to determine whether item order might be systematically influencing DIF. Logistic regression models were re-run for each of the 11 DIF items from the second half of the test. Each of the three variables was entered in a separate step to determine each partial R-square addition. Odds ratios are presented for the full model. In all 11 items the main effect of the disability status grouping variable was significant and for each of those items the odds ratio for the disability status grouping variable was less than 1.0. This strongly demonstrates that students with disabilities under-performed on each of those items relative to students without disabilities when controlling for performance on the first half of the assessment. Similarly, all 11 items had a significant interaction between the disability status grouping variable and the first half ability proxy and the odds ratio for each significant finding was less than 1.0. A significant interaction term with an odds ratio less than 1.0 indicates that a student with disabilities who scored well on the first 27 items would not score as well on the second half of the test as a student without disabilities who had scored similarly on the first 27 items.

**Table 6. State Y Grade 9 Item-level Reading Comprehension Logistic Regression Results for Items Showing DIF with Ability Proxy Based On First 27 Items Score**

Item No.	R-square Results at Each Step in the Sequential Logistic Regression			Odds Ratios–Final Model		
	Step 1 Ability Proxy	Step 2 Ability Proxy and disability status (Uniform)	Step 3 Ability Proxy, disability status and interaction (Non-Uniform)	Ability Proxy	Disability Status	Interaction
32	0.214**	0.216**	0.217**	2.54**	0.53**	0.73 **
37	0.259**	0.259**	0.262**	2.92**	0.56**	0.66**
39	0.277**	0.278**	0.280**	3.08**	0.51**	0.67**
40	0.226**	0.226**	0.230**	2.73**	0.64**	0.62**
41	0.154**	0.154**	0.158**	2.29**	0.80**	0.62**
42	0.105**	0.105**	0.110**	1.95**	0.72**	0.62**
44	0.160**	0.161**	0.163**	2.22**	0.55**	0.68**
48	0.163**	0.168**	0.169**	2.12**	0.49**	0.86**
49	0.214**	0.214**	0.217**	2.72**	0.65**	0.63**
51	0.103**	0.103**	0.106**	1.92**	0.70**	0.68**
54	0.221**	0.224**	0.226**	2.57**	0.46**	0.70**

Note: \* denotes significance at  $p < .05$ . \*\* denotes significance at  $p < .01$

**Word Analysis.** Table 7 presents DIF results from State Y in grade 9 for the 30-item Word Analysis subscale. The total score on the 30 items served as an ability proxy in this model. Items were identified as DIF when the R-square change between steps 1 and 2 was at least 0.003 and was significant at  $p < .01$ . With this model, only one item showed DIF.

In the second model, in which the ability proxy was calculated only from the first 15 items, 12 items showed DIF, 10 of which were located in the second half of the assessment. The effect sizes using the first half ability proxy were larger, especially for the items from the second half of the assessment.

**Table 7. State Y Grade 9 Item-level Word Analysis**

Ability Proxy	Total Number of Items	Number of Items Showing DIF		
		Items 1–15	Items 16–30	All Items
Model 1	30	0	1	1
Model 2	30	2	10	12

*Note:* In Model 1, the total score was used as an ability proxy. In Model 2, the score on the first 15 items was used as an ability proxy.

Items that were found to exhibit DIF from the model in Table 7 were examined more closely in Table 8 to determine whether item order might be systematically influencing DIF. Logistic regression models were re-run for each of the 10 DIF items from the second half of the test. Each of the three variables was entered in a separate step to determine each partial R-square addition. Odds ratios are presented for the full model. In all 10 items the main effect of the disability status grouping variable was significant and for each of those items the odds ratio for the disability status grouping variable was less than 1.0. Again this strongly demonstrates that students with disabilities under-performed on each of those items relative to students without disabilities when controlling for performance on the first half of the assessment. Additionally 8 of these 10 items had a significant interaction between the disability status grouping variable and the first half ability proxy and the odds ratio for each significant finding was less than 1.0.

**Table 8. State Y Grade 9 Item-level Word Analysis Logistic Regression Results for Items Showing DIF with Ability Proxy Based On First 15 Items Score**

Item	R-square results at each step in the sequential logistic regression			Odds Ratios–Final Model		
	Step 1 Ability Proxy	Step 2 Ability Proxy and Disability Status (Uniform)	Step 3 Ability Proxy, Disability Status and Interaction (Non-Uniform)	Ability Proxy	Disability Status	Interaction
18	0.141**	0.145**	0.146**	2.21**	0.56**	0.87**
19	0.120**	0.129**	0.129	2.14**	0.49**	0.98
20	0.030**	0.036**	0.036	1.33**	0.56**	1.02
21	0.203**	0.210**	0.211**	2.78**	0.42**	0.75**
23	0.157**	0.166**	0.166**	2.26**	0.45**	0.91**
24	0.202**	0.204**	0.205**	2.61**	0.58**	0.82**
25	0.204**	0.204**	0.208**	2.80**	0.86**	0.56**
27	0.211**	0.217**	0.218**	2.67**	0.46**	0.79**
28	0.219**	0.224**	0.225**	2.75**	0.47**	0.76**
30	0.247**	0.263**	0.264**	3.31**	0.29**	0.72**

Note: \* denotes significance at  $p < .05$ . \*\* denotes significance at  $p < .01$

### State X Grade 3

**Reading Comprehension.** Table 9 presents DIF results from State X in grade 3 for the 54-item Reading Comprehension subscale. The total score on the 54 items served as an ability proxy in this model. Items were identified as DIF when the R-square change between steps 1 and 2 was at least 0.003 and was significant at  $p < 0.01$ . There were just three items that showed DIF, only one of which was located in the second half of the assessment.

In the second model, in which the ability proxy was calculated only from the first 27 items, seven items showed DIF, five of which were located in the second half of the assessment. The effect sizes using the ability proxy based on the score from the first half were slightly larger for the items from the second half of the assessment.



**Table 9. State X Grade 3 Item-level Reading Comprehension**

Ability Proxy	Total Number of Items	Number of Items Showing DIF		
		Items 1–27	Items 28–54	All Items
Model 1	54	2	1	3
Model 2	54	2	5	7

*Note:* In Model 1, the total score was used as an ability proxy. In Model 2, the score on the first 27 items was used as an ability proxy.

Items that were found to exhibit DIF from Model 2 in Table 9 were examined more closely in Table 10 to determine whether item order might be systematically influencing DIF. Logistic regression models were re-run for each of the five DIF items from the second half of the test. Each of the three variables was entered in a separate step to determine each partial R-square addition. Odds ratios are presented for the full model. In three of the five items the main effect of the disability status grouping variable was significant and for all three of those items the odds ratio for the disability status grouping variable was less than 1.0. This seems to suggest that students with disabilities under-performed on each of those items relative to students without disabilities when controlling for performance on the first half of the assessment. Similarly all five of the DIF items had a significant interaction between the disability status grouping variable and the first half ability proxy and the odds ratio for each significant finding was less than 1.0. A significant interaction term with an odds ratio less than 1.0 indicates that a student with disabilities who scored well on the first 27 items would not score as well on the second half of the test relative to a student without disabilities who had scored similarly on the first 27 items.

There were fewer items exhibiting DIF in grade 3 Reading Comprehension than in grade 9 Reading Comprehension in State X.

**Table 10. State X Grade 3 Item-level Reading Comprehension Logistic Regression Results for Items Showing DIF with Ability Proxy Based On First 27 Items Score**

Item No.	R-square Results at Each Step in the Sequential Logistic Regression			Odds Ratios–Final Model		
	Step 1 Ability Proxy	Step 2 Ability Proxy and disability status (Uniform)	Step 3 Ability Proxy, disability status and interaction (Non-Uniform)	Ability Proxy	Disability Status	Interaction
29	0.256**	0.256	0.259**	3.09**	0.86	0.64 **
31	0.292**	0.294**	0.296**	3.41**	0.50**	0.71**
34	0.317**	0.317	0.320**	3.62**	0.62**	0.64**
42	0.263**	0.264	0.266**	3.04**	0.71**	0.69**
43	0.246**	0.246	0.249**	2.91**	0.84	0.68**

Note: \* denotes significance at  $p < .05$ . \*\* denotes significance at  $p < .01$

**Word Analysis.** Table 11 presents DIF results from State X in grade 3 for the 30-item Word Analysis subscale. The total score on the 30 items served as an ability proxy in this model. Items were identified as DIF when the R-square change between steps 1 and 2 was at least 0.003 and was significant at  $p < 0.01$ . There were seven items that showed DIF, four of which were located in the second half of the assessment (Items 16 through 30).

The second model used a similar method except that the ability proxy was calculated only from the first 15 items. Using this method there were nine items that showed DIF, just two of which came from the second half of the assessment (Items 16 through 30). The number of items showing DIF under this model from the second half of the test decreased.

**Table 11. State X Grade 3 Item-level Word Analysis**

Ability Proxy	Total Number of Items	Number of Items Showing DIF		
		Items–15	Items 16–30	All Items
Model 1	30	3	4	7
Model 2	30	7	2	9

Note: In Model 1, the total score was used as an ability proxy. In Model 2, the score on the first 15 items was used as an ability proxy.

These findings are unlike those seen in grade 9. In grade 9 there were more items exhibiting DIF from the second half than from the first half of the test when the score on the first half of the test was used as the ability proxy. For grade 3, logistic regression models were re-run for the two DIF items from the second half of the test, and are presented in Table 12. Only one of the two items had a significant main effect of the disability status grouping variable, and the odds ratio was less than 1.0. These results suggest that in State X the factors influencing the results for students with disabilities in grade 9 in WA were not present for students with disabilities in grade 3.

**Table 12. State X Grade 3 Item-level Word Analysis Logistic Regression Results for Items Showing DIF with Ability Proxy Based On First 15 Items Score**

Item	R-square Results at Each Step in the Sequential Logistic Regression			Odds Ratios--Final Model		
	Step 1 Ability Proxy	Step 2 Ability Proxy and Disability Status (Uniform)	Step 3 Ability Proxy, Disability Status and Interaction (Non-Uniform)	Ability Proxy	Disability Status	Interaction
16	0.342**	0.345**	0.345	4.05**	0.64	1.04
25	0.137**	0.140*	0.141	1.95**	0.70**	1.20

Note: \* denotes significance at  $p < .05$ . \*\* denotes significance at  $p < .01$

### State Y Grade 3

**Reading Comprehension.** Table 13 presents DIF results from State Y in grade 3 for the 54-item Reading Comprehension subscale. The total score on the 54 items served as an ability proxy in this model. Items were identified as DIF when the R-square change between steps 1 and 2 was at least 0.003 and was significant at  $p < 0.01$ . There were no items that showed DIF in grade 3 using this method.

In the second model, in which the ability proxy was calculated only from the first 27 items, 7 items showed DIF, one of which was located in the second half of the assessment. The number of items showing DIF under this model from the second half of the test decreased.

**Table 13. State Y Grade 3 Item-level Reading Comprehension**

Ability Proxy	Total Number of Items	Number of Items Showing DIF		
		Items 1–27	Items 28–54	All Items
Model 1	54	0	0	0
Model 2	54	6	1	7

*Note:* In Model 1, the total score was used as an ability proxy. In Model 2, the score on the first 27 items was used as an ability proxy.

These findings are unlike those seen in grade 9. In grade 9 there were more items exhibiting DIF from the second half than from the first half of the test when the score on the first half of the test was used as the ability proxy. For grade 3, logistic regression models were re-run for the one item showing DIF from the second half of the test, and are presented in Table 14. There was a significant main effect of the disability status variable and the odds ratio was less than 1.0. These results suggest that in State Y the factors influencing the results for students with disabilities in grade 9 in RC were not present for students with disabilities in grade 3.

**Table 14. State Y Grade 3 Item-level Reading Comprehension Logistic Regression Results for Items Showing DIF with Ability Proxy Based On First 27 Items Score**

Item No.	R-square results at each step in the sequential logistic regression			Odds Ratios–Final Model		
	Step 1 Ability Proxy	Step 2 Ability Proxy and disability status (Uniform)	Step 3 Ability Proxy, disability status and interaction (Non-Uniform)	Ability Proxy	Disability Status	Interaction
45	0.233**	0.233**	0.236**	2.80**	0.88**	0.66**

*Note:* \* denotes significance at  $p < .05$ . \*\* denotes significance at  $p < .01$

**Word Analysis.** Table 15 presents DIF results from State Y in grade 3 for the 30-item Word Analysis subscale. The total score on the 30 items served as an ability proxy in this model. Items were identified as DIF when the R-square change between steps 1 and 2 was at least 0.003 and was significant at  $p < .01$ . There were no items that showed DIF using the total score on the 30 items as an ability proxy.

In the second model, in which the ability proxy was calculated only from the first 15 items, 12 items showed DIF, 4 of which were located in the second half of the assessment.

**Table 15. State Y Grade 3 Item-level Word Analysis**

Ability Proxy	Total Number of Items	Number of Items Showing DIF		
		Items 1–15	Items 16–30	All Items
Model 1	30	0	0	0
Model 2	30	8	4	12

*Note:* In Model 1, the total score was used as an ability proxy. In Model 2, the score on the first 15 items was used as an ability proxy.

These findings are unlike those seen in grade 9. In grade 9 there were more items exhibiting DIF from the second half than from the first half of the test when the score on the first half of the test was used as the ability proxy. Logistic regression models were re-run for the four items that did indicate DIF from the second half of the test, and are presented in Table 16. The odds ratio for each of these items indicates that students with disabilities under-performed relative to students without disabilities after controlling for performance on the first half of the test.

**Table 16. State Y Grade 3 Item-level Word Analysis Logistic Regression Results for Items Showing DIF with Ability Proxy Based On First 15 Items Score**

Item	R-square Results at Each Step in the Sequential Logistic Regression			Odds Ratios–Final Model		
	Step 1 Ability Proxy	Step 2 Ability Proxy and Disability Status (Uniform)	Step 3 Ability Proxy, Disability Status and Interaction (Non-Uniform)	Ability Proxy	Disability Status	Interaction
16	0.412**	0.414**	0.415**	4.85**	0.50**	0.77**
18	0.470**	0.476**	0.478**	7.68**	0.24**	0.60**
25	0.173**	0.176**	0.176	2.21**	0.69**	1.00
30	0.307**	0.307**	0.310**	3.46**	0.68**	0.65**

*Note:* \* denotes significance at  $p < .05$ . \*\* denotes significance at  $p < .01$

## Discussion

Students with disabilities tend to perform at lower levels than students without disabilities. While their lower performance can be attributed to their specific disability, there may be other factors that potentially interfere with their performance. It is necessary to identify such factors and reduce their interference, so that we may obtain accurate measurements of the knowledge of students with disabilities. Recent reauthorizations of federal legislations render it imperative that the instruction and assessment of students with disabilities are as fair and adequate as possible. While we recognize that factors related to instruction and assessment are intricately intertwined, only a relatively small portion of students with disabilities have conditions that lower their performance potential. This study does not address that issue but instead focuses specifically on factors related directly to the assessments and the accuracy as to which they reflect what students learn. The present study explored whether items in a reading assessment functioned differentially for students with disabilities, as compared to their peers without disabilities. Results of this study can provide insight into potential factors affecting the accessibility of reading assessments for students with disabilities, as part of an effort to improve assessments for all students.

The following research questions guided this study:

1. Do items on standardized Reading Comprehension (RC) and Word Analysis (WA) subscales exhibit Differential Item Functioning (DIF) for students with disabilities?
2. Are more items that exhibit DIF for students with disabilities located in the second half of RC and WA subscales rather than in the first half?
3. Do students with disabilities consistently under-perform on items located in the second half relative to items located in the first half, as compared to students without disabilities?
4. Do the results of DIF vary by grade (grade 3 and grade 9)?

To answer these research questions, student responses on multiple-choice items were compared across the disability status categories in two reading subscales of the Stanford 9, Reading Comprehension and Word Analysis, in two grade levels (3 and 9) from public schools in two different states (State X and State Y). A multi-step logistic regression procedure was used. Because it is essential in DIF analysis that the two groups being compared are matched on ability level, ability proxies were used based on either the total score of the subscale, or the combined score on the first half of the subscale.

After accounting for reading ability, results for grade 9 in both states indicated that there are a number of items that exhibit DIF for students with disabilities on both the RC and

WA subscales. Results also indicated that the items exhibiting DIF for students with disabilities are more likely to be located in the second half of the RC and WA subscales. When the reading ability proxy was based on the combined score from the first half of the RC or WA subscales, the effect size for DIF increased for the items located in the second half. Furthermore, students with disabilities consistently under-performed on the second half of the items relative to the first half of the items.

These results were not seen for grade 3. In other words, there were fewer items that were shown to exhibit DIF for students with disabilities than what was found in grade 9. This was true for both the RC and WA subscales and for both states. In grade 3, items that were shown to exhibit DIF for students with disabilities were no more likely to be located in the second half of the assessments than they were in the first half of the assessments.

The findings of this study have multiple implications. There are differences between grade 3 and grade 9, that may result from cognitive development of reading skills, or perhaps the differences in assessment standards for those grades, or that students with disabilities are more clearly identified as having disabilities in older years. In grade 9, we might speculate over what factors contribute to the diminishing performance for students with disabilities as the test progresses. Perhaps students with disabilities did not have sufficient time or energy to complete the test and rushed through the answers at the end. It could be that they reached a certain cognitive overload, lost motivation, or became fatigued or frustrated. Our companion report, that examines differential *distractor* functioning, found that students with disabilities in grade 9, appear to be making more random guesses rather than “educated” guesses in items located in the second half of the assessments, as compared to their non-disabled peers (see Abedi, Leon, & Kao, 2006, for more detail). More research is needed to determine the actual cause or causes. Qualitative research with students may potentially shed some light on these factors.

This study has several major limitations. For instance, it does not differentiate between categories of disabilities. Students with disabilities are not a homogeneous subgroup. Not only are there different types of disabilities, but even within the same type of disability there are differences among individuals. Further insight could be gained from analyzing data by specific disability groups. This study was also limited in terms of scope. We did not have access to information on testing accommodations. Although our study was conducted assuming that students were properly accommodated, we do not know this for sure. It could be that students with disabilities did not receive adequate or appropriate accommodations, and knowing this could inform the results. Also, we did not have access to the actual test booklets or test items, which could provide further insight into the findings. Future studies should take into account accommodations and examine test booklets.

Nevertheless, findings of this study provide evidence that other factors related to the assessments may contribute to the performance gap between students with disabilities and

their peers without disabilities. Controlling for factors that are not related to the content being assessed may help test developers provide more accessible and more valid assessments for students with disabilities. Additionally, being cognizant that other factors exist may help when interpreting test results for students with disabilities, especially in the context of accountability.



## References

- Abedi, J., Leon, S., & Kao, J. (2006). *Examining differential distractor functioning in reading assessments for students with disabilities (PARA report)*. Minneapolis, MN: Partnership for Accessible Reading Assessment.
- Bolt, S. E. (2004). *Using DIF analyses to examine several commonly-held beliefs about testing accommodations for students with disabilities*. Paper presented at the Annual Meeting of the National Council on Measurement in Education, San Diego, CA. Retrieved June 20, 2006, from <http://education.umn.edu/NCEO/Presentations/NCME04bolt.pdf>
- Clauser, B. E., & Mazor, K. M. (1998). Using statistical procedures to identify differentially functioning test items. *Educational Measurement: Issues and Practice*, 17(1), 31–44.
- Cohen, A. S., Gregg, N., & Deng, M. (2005). The role of extended time and item content on a high-stakes mathematics test. *Learning Disabilities Research & Practice*, 20(4), 225–233.
- Haladyna, T. M., & Downing, S. M. (2004). Construct-irrelevant variance in high-stakes testing. *Educational Measurement: Issues and Practice*, 23(1), 17–27.
- HarcourtAssessment.com. (n.d.). *Stanford Achievement Test series, Ninth edition – Complete battery*. Retrieved May 18, 2006, from <http://harcourtassessment.com/hai/Product-LongDesc.aspx?ISBN=E132C&Catalog=TPC-USCatalog&Category=AchievementAccountability>
- Hauser, C., & Kingsbury, G. (2004). *Differential item functioning and differential test functioning in the “Idaho Standards Achievement Tests” for spring 2003*. Lake Oswego, OR: Northwest Evaluation Association.
- Klein, J. A., Wiley, H. I., & Thurlow, M. L. (2006). *Uneven transparency: NCLB tests take precedence in public assessment reporting for students with disabilities (Technical Report 43)*. Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes.
- Koretz, D., & Hamilton, L. (1999). *Assessing students with disabilities in Kentucky: The effects of accommodations, format, and subject*. Los Angeles, CA: University of California, National Center for Research on Evaluation, Standards, and Student Testing. Retrieved June 28, 2006, from <http://www.cresst.org/Reports/TECH498.pdf>
- No Child Left Behind Act of 2001*, Pub. L. No. 107–110, 115 Stat. 1425 (2002).
- Penfield, R. D., & Lam, T. C. M. (2000). Assessing differential item functioning in performance assessment. *Educational Measurement: Issues and Practice*, 19(3), 5–15.

Snetzler, S., & Qualls, A. L. (2000). Examination of differential item functioning on a standardized achievement battery with limited English proficient students. *Educational and Psychological Measurement*, 60(4), 564–577.

Thurlow, M. L., Moen, R. E., & Wiley, H. I. (2005). *Annual performance reports: 2002–2003 state assessment data*. Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes. Retrieved May 2, 2006, from <http://education.umn.edu/nceo/OnlinePubs/APRsummary2005.pdf>

U.S. General Accounting Office (2005). *No Child Left Behind Act: Most students with disabilities participated in statewide assessments, but inclusion options could be improved*. Washington, DC: Author.

Zenisky, A. L., Hambleton, R. K., & Robin, F. (2004). DIF detection and interpretation in large-scale science assessments: Informing item writing practices. *Educational Assessment*, 9(1–2), 61–78.

Zumbo, B. D. (1999). *A handbook on the theory and methods of differential item functioning (DIF): Logistic regression modeling as a unitary framework for binary and Likert-type (ordinal) item scores*. Ottawa, ON: Directorate of Human Resources Research and Evaluation, Department of National Defense. Retrieved March 10, 2006, from <http://educ.ubc.ca/faculty/zumbo/DIF/handbook.pdf>

## Acknowledgments

The work reported here was supported under a grant from the U.S. Department of Education, Office of Special Education Programs. The findings and opinions expressed in this report are those of the authors and do not necessarily reflect the positions or policies of the Office of Special Education Programs or the U.S. Department of Education.

The authors acknowledge the valuable contribution of colleagues in this study. The authors are thankful to Martha Thurlow, Ross Moen, Christopher Johnstone, and other staff at the National Center on Educational Outcomes, and other members of the Partnership for Accessible Reading Assessment for their helpful comments and suggestions. Danna Schacter at CRESST also helped us substantially with the preparation of this manuscript. The authors are also grateful to Eva Baker for her support of this work, and to Joan Herman for her extensive involvement, advice, and support of this work.



## Appendix

### Detailed DIF Results

**Table A1. State X Grade 9 Item-level Reading Comprehension Ability Proxy Based On All 54 Items**

Item	R-squared values at each step in the sequential hierarchical regression		DIF results	
	Step 1 Ability Proxy	Step 2 Ability Proxy, Disability Status, Interaction	Chi-Sq P-value	Change in R-Square (Effect size)
1	0.195	0.195	0.388	0.000
2	0.276	0.277	0.045	0.001
3	0.329	0.332	0.001	0.003
4	0.269	0.270	0.101	0.001
5	0.357	0.359	0.081	0.001
6	0.290	0.292	0.020	0.001
7	0.161	0.162	0.167	0.001
8	0.240	0.244	0.000	0.004
9	0.314	0.314	0.264	0.000
10	0.300	0.301	0.405	0.001
11	0.385	0.385	0.708	0.000
12	0.167	0.170	0.003	0.003
13	0.216	0.218	0.021	0.002
14	0.387	0.388	0.125	0.001
15	0.170	0.171	0.072	0.001
16	0.166	0.167	0.761	0.001
17	0.192	0.193	0.135	0.001
18	0.223	0.224	0.253	0.001
19	0.264	0.266	0.012	0.002
20	0.210	0.212	0.014	0.002
21	0.311	0.311	0.321	0.000
22	0.166	0.166	0.720	0.000
23	0.205	0.207	0.013	0.002
24	0.212	0.217	0.000	0.005
25	0.221	0.222	0.214	0.001
26	0.220	0.221	0.242	0.001
27	0.289	0.289	0.854	0.000
28	0.125	0.125	0.671	0.000
29	0.256	0.258	0.006	0.002

Item	R-squared values at each step in the sequential hierarchical regression		DIF results	
	Step 1 Ability Proxy	Step 2 Ability Proxy, Disability Status, Interaction	Chi-Sq P-value	Change in R-Square (Effect size)
30	0.173	0.174	0.514	0.001
31	0.328	0.328	0.940	0.000
32	0.268	0.269	0.072	0.001
33	0.283	0.286	0.000	0.003
34	0.293	0.293	0.963	0.000
35	0.077	0.080	0.003	0.003
36	0.117	0.120	0.001	0.003
37	0.336	0.339	0.000	0.003
38	0.046	0.046	0.271	0.000
39	0.382	0.383	0.115	0.001
40	0.325	0.327	0.007	0.002
41	0.194	0.201	0.000	0.007
42	0.173	0.182	0.000	0.009
43	0.375	0.375	0.368	0.000
44	0.261	0.264	0.003	0.003
45	0.216	0.217	0.448	0.001
46	0.322	0.323	0.052	0.001
47	0.178	0.181	0.001	0.003
48	0.227	0.230	0.000	0.003
49	0.323	0.329	0.000	0.006
50	0.196	0.199	0.001	0.003
51	0.171	0.175	0.000	0.004
52	0.214	0.223	0.000	0.009
53	0.095	0.096	0.365	0.001
54	0.347	0.349	0.003	0.002

**Table A2. State X Grade 9 Item-level Reading Comprehension Ability Proxy Based On First 27 Items**

Item	R-squared values at each step in the sequential hierarchical regression		DIF results	
	Step 1 Ability Proxy	Step 2 Ability Proxy, Disability Status, Interaction	Chi-Sq P-value	Change in R-Square (Effect size)
1	0.242	0.242	0.915	0.000
2	0.311	0.314	0.000	0.002
3	0.383	0.388	0.000	0.005
4	0.314	0.315	0.228	0.001
5	0.403	0.403	0.294	0.001
6	0.288	0.292	0.001	0.004
7	0.205	0.207	0.005	0.002
8	0.246	0.250	0.000	0.004
9	0.364	0.364	0.602	0.000
10	0.341	0.342	0.024	0.001
11	0.407	0.408	0.735	0.001
12	0.213	0.213	0.142	0.000
13	0.239	0.239	0.227	0.000
14	0.414	0.416	0.006	0.002
15	0.197	0.199	0.006	0.002
16	0.192	0.192	0.954	0.000
17	0.217	0.219	0.006	0.002
18	0.245	0.246	0.089	0.001
19	0.291	0.293	0.011	0.002
20	0.228	0.234	0.000	0.006
21	0.325	0.325	0.422	0.000
22	0.192	0.194	0.076	0.002
23	0.225	0.231	0.000	0.006
24	0.220	0.228	0.000	0.008
25	0.238	0.239	0.043	0.001
26	0.235	0.235	0.667	0.000
27	0.306	0.307	0.206	0.000
28	0.098	0.100	0.017	0.002
29	0.176	0.179	0.001	0.003
30	0.118	0.120	0.042	0.002
31	0.244	0.246	0.005	0.002
32	0.171	0.179	0.000	0.008

Item	R-squared values at each step in the sequential hierarchical regression		DIF results	
	Step 1 Ability Proxy	Step 2 Ability Proxy, Disability Status, Interaction	Chi-Sq P-value	Change in R-Square (Effect size)
33	0.195	0.201	0.000	0.006
34	0.219	0.221	0.015	0.002
35	0.034	0.037	0.005	0.003
36	0.065	0.073	0.001	0.008
37	0.222	0.225	0.001	0.003
38	0.018	0.018	0.730	0.000
39	0.269	0.271	0.006	0.002
40	0.226	0.229	0.002	0.003
41	0.114	0.121	0.000	0.007
42	0.098	0.106	0.000	0.008
43	0.250	0.254	0.000	0.004
44	0.161	0.168	0.000	0.007
45	0.140	0.144	0.000	0.004
46	0.223	0.225	0.040	0.002
47	0.096	0.098	0.040	0.002
48	0.144	0.155	0.000	0.011
49	0.209	0.218	0.000	0.009
50	0.111	0.113	0.004	0.002
51	0.101	0.104	0.001	0.003
52	0.127	0.132	0.000	0.005
53	0.049	0.049	0.548	0.000
54	0.230	0.240	0.000	0.010



**Table A3. State X Grade 9 Item-level Word Analysis Ability Proxy Based On All 30 Items**

Item	R-squared values at each step in the sequential hierarchical regression		DIF results	
	Step 1 Ability Proxy	Step 2 Ability Proxy, Disability Status, Interaction	Chi-Sq P-value	Change in R-Square (Effect size)
1	0.225	0.227	0.023	0.002
2	0.299	0.301	0.002	0.002
3	0.114	0.114	0.261	0.000
4	0.094	0.099	0.000	0.005
5	0.139	0.147	0.000	0.008
6	0.093	0.094	0.477	0.001
7	0.102	0.106	0.000	0.004
8	0.105	0.106	0.174	0.001
9	0.042	0.044	0.017	0.002
10	0.277	0.277	0.724	0.000
11	0.312	0.312	0.670	0.000
12	0.224	0.225	0.551	0.000
13	0.210	0.211	0.043	0.000
14	0.372	0.375	0.000	0.004
15	0.416	0.418	0.008	0.002
16	0.272	0.273	0.441	0.001
17	0.174	0.175	0.100	0.001
18	0.257	0.260	0.001	0.003
19	0.234	0.239	0.000	0.005
20	0.100	0.104	0.000	0.004
21	0.351	0.352	0.074	0.001
22	0.141	0.141	0.968	0.000
23	0.311	0.313	0.025	0.002
24	0.344	0.347	0.001	0.003
25	0.329	0.343	0.000	0.014
26	0.337	0.342	0.000	0.005
27	0.383	0.385	0.003	0.002
28	0.332	0.334	0.006	0.002
29	0.207	0.212	0.000	0.005
30	0.437	0.440	0.000	0.003

**Table A4. State X Grade 9 Item-level Word Analysis Ability Proxy Based On First 15 Items**

Item	R-squared values at each step in the sequential hierarchical regression		DIF results	
	Step 1 Ability Proxy	Step 2 Ability Proxy, Disability Status, Interaction	Chi-Sq P-value	Change in R-Square (Effect size)
1	0.345	0.346	0.077	0.001
2	0.402	0.405	0.001	0.003
3	0.180	0.181	0.245	0.000
4	0.144	0.149	0.000	0.005
5	0.182	0.190	0.000	0.008
6	0.149	0.150	0.461	0.000
7	0.169	0.176	0.000	0.007
8	0.154	0.155	0.252	0.001
9	0.079	0.083	0.002	0.004
10	0.303	0.303	0.720	0.000
11	0.341	0.343	0.032	0.002
12	0.257	0.257	0.435	0.000
13	0.240	0.240	0.222	0.000
14	0.390	0.393	0.001	0.003
15	0.430	0.436	0.000	0.006
16	0.145	0.151	0.000	0.006
17	0.071	0.075	0.000	0.004
18	0.129	0.147	0.000	0.018
19	0.105	0.125	0.000	0.020
20	0.029	0.043	0.000	0.014
21	0.196	0.208	0.000	0.012
22	0.049	0.053	0.000	0.004
23	0.153	0.169	0.000	0.016
24	0.194	0.199	0.000	0.005
25	0.201	0.209	0.000	0.008
26	0.186	0.188	0.005	0.002
27	0.201	0.215	0.000	0.014
28	0.180	0.190	0.000	0.010
29	0.093	0.094	0.052	0.001
30	0.266	0.290	0.000	0.024

**Table A5. State Y Grade 9 Item-level Reading Comprehension Ability Proxy Based On All 54 Items**

Item	R-squared values at each step in the sequential hierarchical regression		DIF results	
	Step 1 Ability Proxy	Step 2 Ability Proxy, Disability Status, Interaction	Chi-Sq P-value	Change in R-Square (Effect size)
1	0.208	0.208	0.000	0.000
2	0.295	0.295	0.000	0.000
3	0.310	0.310	0.001	0.000
4	0.217	0.317	0.000	0.000
5	0.330	0.331	0.000	0.001
6	0.267	0.267	0.052	0.000
7	0.180	0.180	0.000	0.000
8	0.231	0.231	0.000	0.000
9	0.333	0.334	0.000	0.000
10	0.285	0.285	0.000	0.000
11	0.347	0.347	0.681	0.000
12	0.185	0.185	0.004	0.000
13	0.183	0.183	0.000	0.000
14	0.404	0.404	0.000	0.000
15	0.161	0.161	0.221	0.001
16	0.194	0.195	0.000	0.001
17	0.145	0.145	0.000	0.000
18	0.226	0.226	0.000	0.000
19	0.278	0.279	0.000	0.001
20	0.250	0.250	0.000	0.000
21	0.279	0.280	0.001	0.000
22	0.227	0.227	0.000	0.000
23	0.188	0.188	0.000	0.000
24	0.241	0.241	0.000	0.000
25	0.249	0.249	0.001	0.001
26	0.182	0.182	0.000	0.001
27	0.282	0.282	0.854	0.000
28	0.198	0.198	0.000	0.000
29	0.274	0.275	0.000	0.001
30	0.159	0.160	0.000	0.001
31	0.353	0.353	0.027	0.000
32	0.292	0.292	0.000	0.000

Item	R-squared values at each step in the sequential hierarchical regression		DIF results	
	Step 1 Ability Proxy	Step 2 Ability Proxy, Disability Status, Interaction	Chi-Sq P-value	Change in R-Square (Effect size)
33	0.325	0.325	0.000	0.000
34	0.277	0.277	0.000	0.000
35	0.095	0.095	0.000	0.000
36	0.139	0.139	0.000	0.000
37	0.377	0.378	0.000	0.001
38	0.045	0.045	0.031	0.000
39	0.398	0.398	0.000	0.000
40	0.336	0.338	0.000	0.002
41	0.252	0.254	0.000	0.002
42	0.194	0.196	0.000	0.002
43	0.419	0.419	0.000	0.000
44	0.267	0.268	0.000	0.001
45	0.261	0.261	0.000	0.000
46	0.315	0.315	0.082	0.000
47	0.202	0.202	0.000	0.000
48	0.275	0.275	0.005	0.000
49	0.340	0.342	0.000	0.002
50	0.230	0.230	0.000	0.000
51	0.187	0.189	0.000	0.002
52	0.251	0.253	0.000	0.002
53	0.121	0.121	0.000	0.000
54	0.356	0.357	0.000	0.001

**Table A6. State Y Grade 9 Item-level Reading Comprehension Ability Proxy Based On First 27 Items**

Item	R-squared values at each step in the sequential hierarchical regression		DIF results	
	Step 1 Ability Proxy	Step 2 Ability Proxy, Disability Status, Interaction	Chi-Sq P-value	Change in R-Square (Effect size)
1	0.239	0.239	0.026	0.000
2	0.337	0.338	0.000	0.001
3	0.354	0.355	0.000	0.001
4	0.360	0.360	0.000	0.000
5	0.375	0.375	0.031	0.000
6	0.293	0.294	0.000	0.001
7	0.215	0.217	0.000	0.002
8	0.262	0.264	0.000	0.002
9	0.380	0.380	0.000	0.000
10	0.324	0.325	0.000	0.001
11	0.380	0.380	0.000	0.000
12	0.227	0.227	0.000	0.000
13	0.212	0.213	0.000	0.001
14	0.434	0.435	0.000	0.001
15	0.188	0.189	0.000	0.001
16	0.223	0.225	0.000	0.002
17	0.168	0.168	0.000	0.000
18	0.252	0.253	0.000	0.001
19	0.296	0.296	0.537	0.000
20	0.255	0.258	0.000	0.003
21	0.297	0.297	0.000	0.000
22	0.248	0.249	0.000	0.001
23	0.212	0.214	0.000	0.002
24	0.264	0.268	0.000	0.004
25	0.266	0.267	0.000	0.001
26	0.203	0.203	0.068	0.000
27	0.304	0.304	0.000	0.000
28	0.131	0.133	0.000	0.002
29	0.194	0.195	0.000	0.001
30	0.107	0.109	0.000	0.002
31	0.267	0.269	0.000	0.002
32	0.214	0.217	0.000	0.003

Item	R-squared values at each step in the sequential hierarchical regression		DIF results	
	Step 1 Ability Proxy	Step 2 Ability Proxy, Disability Status, Interaction	Chi-Sq P-value	Change in R-Square (Effect size)
33	0.239	0.241	0.000	0.002
34	0.201	0.203	0.000	0.000
35	0.048	0.049	0.000	0.001
36	0.085	0.087	0.000	0.002
37	0.259	0.262	0.000	0.003
38	0.018	0.018	0.000	0.000
39	0.277	0.280	0.000	0.003
40	0.226	0.230	0.000	0.004
41	0.154	0.158	0.000	0.004
42	0.105	0.110	0.000	0.005
43	0.279	0.281	0.000	0.002
44	0.160	0.163	0.000	0.003
45	0.163	0.164	0.000	0.001
46	0.188	0.190	0.000	0.002
47	0.107	0.109	0.000	0.002
48	0.163	0.169	0.000	0.006
49	0.214	0.217	0.000	0.003
50	0.125	0.127	0.000	0.002
51	0.103	0.106	0.000	0.003
52	0.148	0.150	0.000	0.002
53	0.055	0.056	0.000	0.001
54	0.221	0.226	0.000	0.005

**Table A7. State Y Grade 9 Item-level Word Analysis Ability Proxy Based On All 30 Items**

Item	R-squared values at each step in the sequential hierarchical regression		DIF results	
	Step 1 Ability Proxy	Step 2 Ability Proxy, Disability Status, Interaction	Chi-Sq P-value	Change in R-Square (Effect size)
1	0.272	0.272	0.177	0.000
2	0.392	0.392	0.001	0.000
3	0.140	0.140	0.001	0.000
4	0.140	0.141	0.000	0.001
5	0.124	0.125	0.000	0.001
6	0.090	0.090	0.000	0.000
7	0.131	0.133	0.000	0.002
8	0.108	0.108	0.000	0.000
9	0.053	0.053	0.013	0.000
10	0.293	0.294	0.000	0.000
11	0.315	0.315	0.000	0.000
12	0.234	0.234	0.000	0.000
13	0.219	0.219	0.735	0.000
14	0.414	0.415	0.000	0.001
15	0.413	0.414	0.000	0.001
16	0.243	0.244	0.000	0.001
17	0.147	0.147	0.000	0.000
18	0.269	0.269	0.101	0.000
19	0.252	0.252	0.000	0.000
20	0.093	0.093	0.001	0.000
21	0.370	0.370	0.000	0.000
22	0.129	0.129	0.166	0.000
23	0.311	0.311	0.002	0.000
24	0.350	0.350	0.000	0.000
25	0.341	0.346	0.000	0.005
26	0.352	0.353	0.000	0.001
27	0.387	0.387	0.000	0.000
28	0.382	0.383	0.000	0.001
29	0.272	0.273	0.000	0.001
30	0.425	0.426	0.000	0.001

**Table A8. State Y Grade 9 Item-level Word Analysis Ability Proxy Based On First 15 Items**

Item	R-squared values at each step in the sequential hierarchical regression		DIF results	
	Step 1 Ability Proxy	Step 2 Ability Proxy, Disability Status, Interaction	Chi-Sq P-value	Change in R-Square (Effect size)
1	0.306	0.307	0.000	0.001
2	0.396	0.398	0.000	0.002
3	0.207	0.207	0.321	0.000
4	0.195	0.198	0.000	0.003
5	0.180	0.182	0.000	0.002
6	0.150	0.150	0.186	0.000
7	0.190	0.195	0.000	0.005
8	0.163	0.165	0.000	0.002
9	0.100	0.102	0.000	0.002
10	0.323	0.324	0.720	0.000
11	0.336	0.336	0.032	0.002
12	0.272	0.272	0.000	0.000
13	0.260	0.260	0.000	0.000
14	0.426	0.427	0.000	0.001
15	0.424	0.426	0.000	0.002
16	0.145	0.147	0.000	0.002
17	0.065	0.066	0.000	0.001
18	0.141	0.146	0.000	0.005
19	0.120	0.129	0.000	0.009
20	0.030	0.036	0.000	0.006
21	0.203	0.211	0.000	0.008
22	0.047	0.049	0.000	0.002
23	0.157	0.166	0.000	0.009
24	0.202	0.205	0.000	0.003
25	0.204	0.208	0.000	0.008
26	0.202	0.203	0.000	0.001
27	0.211	0.218	0.000	0.007
28	0.219	0.225	0.000	0.006
29	0.146	0.148	0.000	0.002
30	0.247	0.264	0.000	0.017



**Table A9. State X Grade 3 Item-level Reading Comprehension Ability Proxy Based On All 54 Items**

Item	R-squared values at each step in the sequential hierarchical regression		DIF results	
	Step 1 Ability Proxy	Step 2 Ability Proxy, Disability Status, Interaction	Chi-Sq P-value	Change in R-Square (Effect size)
1	0.337	0.338	0.050	0.001
2	0.465	0.465	0.448	0.000
3	0.326	0.328	0.018	0.002
4	0.201	0.202	0.172	0.001
5	0.390	0.392	0.036	0.002
6	0.147	0.148	0.048	0.001
7	0.278	0.279	0.131	0.001
8	0.074	0.074	0.561	0.000
9	0.092	0.092	0.769	0.000
10	0.290	0.291	0.036	0.001
11	0.295	0.298	0.000	0.003
12	0.347	0.348	0.198	0.001
13	0.239	0.239	0.147	0.001
14	0.217	0.217	0.885	0.000
15	0.063	0.064	0.789	0.001
16	0.299	0.299	0.943	0.000
17	0.343	0.344	0.331	0.001
18	0.040	0.043	0.001	0.003
19	0.178	0.178	0.544	0.000
20	0.266	0.267	0.098	0.001
21	0.393	0.393	0.882	0.000
22	0.220	0.221	0.478	0.001
23	0.284	0.285	0.739	0.001
24	0.352	0.352	0.339	0.000
25	0.482	0.483	0.312	0.001
26	0.366	0.366	0.511	0.000
27	0.295	0.296	0.161	0.001
28	0.249	0.250	0.062	0.001
29	0.331	0.332	0.019	0.001
30	0.314	0.314	0.478	0.000
31	0.414	0.414	0.083	0.000
32	0.357	0.357	0.821	0.000

Item	R-squared values at each step in the sequential hierarchical regression		DIF results	
	Step 1 Ability Proxy	Step 2 Ability Proxy, Disability Status, Interaction	Chi-Sq P-value	Change in R-Square (Effect size)
33	0.191	0.192	0.232	0.001
34	0.470	0.471	0.039	0.001
35	0.375	0.376	0.303	0.001
36	0.475	0.475	0.390	0.001
37	0.384	0.385	0.015	0.001
38	0.216	0.218	0.003	0.002
39	0.238	0.238	0.627	0.000
40	0.387	0.388	0.208	0.001
41	0.367	0.368	0.202	0.001
42	0.416	0.417	0.015	0.001
43	0.406	0.408	0.000	0.002
44	0.243	0.245	0.011	0.002
45	0.401	0.404	0.000	0.003
46	0.268	0.269	0.125	0.001
47	0.283	0.285	0.005	0.002
48	0.241	0.243	0.023	0.002
49	0.371	0.373	0.017	0.002
50	0.257	0.257	0.885	0.000
51	0.135	0.135	0.213	0.000
52	0.379	0.380	0.012	0.001
53	0.185	0.185	0.581	0.000
54	0.171	0.171	0.082	0.000

**Table A10. State X Grade 3 Item-level Reading Comprehension Ability Proxy Based On First 27 Items**

Item	R-squared values at each step in the sequential hierarchical regression		DIF results	
	Step 1 Ability Proxy	Step 2 Ability Proxy, Disability Status, Interaction	Chi-Sq P-value	Change in R-Square (Effect size)
1	0.403	0.404	0.102	0.001
2	0.574	0.574	0.917	0.000
3	0.417	0.418	0.032	0.001
4	0.244	0.244	0.475	0.001
5	0.486	0.487	0.345	0.001
6	0.196	0.197	0.111	0.001
7	0.317	0.317	0.273	0.000
8	0.130	0.130	0.162	0.000
9	0.136	0.136	0.892	0.000
10	0.318	0.322	0.000	0.004
11	0.345	0.347	0.001	0.002
12	0.390	0.391	0.107	0.001
13	0.294	0.295	0.277	0.001
14	0.248	0.248	0.542	0.000
15	0.086	0.087	0.314	0.001
16	0.315	0.316	0.036	0.001
17	0.379	0.379	0.845	0.000
18	0.049	0.053	0.000	0.004
19	0.200	0.200	0.977	0.000
20	0.300	0.301	0.006	0.001
21	0.420	0.421	0.681	0.001
22	0.249	0.250	0.039	0.001
23	0.329	0.330	0.323	0.001
24	0.380	0.381	0.005	0.001
25	0.471	0.473	0.003	0.002
26	0.365	0.366	0.018	0.001
27	0.310	0.311	0.013	0.001
28	0.195	0.196	0.054	0.001
29	0.256	0.259	0.003	0.003
30	0.231	0.232	0.042	0.001
31	0.292	0.296	0.000	0.004
32	0.271	0.272	0.073	0.001

Item	R-squared values at each step in the sequential hierarchical regression		DIF results	
	Step 1 Ability Proxy	Step 2 Ability Proxy, Disability Status, Interaction	Chi-Sq P-value	Change in R-Square (Effect size)
33	0.115	0.115	0.122	0.000
34	0.317	0.320	0.000	0.003
35	0.269	0.270	0.065	0.001
36	0.334	0.336	0.002	0.002
37	0.252	0.254	0.003	0.002
38	0.128	0.130	0.031	0.002
39	0.140	0.140	0.192	0.000
40	0.247	0.249	0.023	0.002
41	0.238	0.238	0.374	0.000
42	0.263	0.266	0.000	0.003
43	0.246	0.249	0.001	0.003
44	0.129	0.131	0.012	0.002
45	0.238	0.240	0.001	0.002
46	0.153	0.153	0.884	0.000
47	0.138	0.140	0.005	0.002
48	0.206	0.207	0.041	0.001
49	0.198	0.199	0.008	0.001
50	0.142	0.143	0.132	0.001
51	0.069	0.070	0.014	0.001
52	0.213	0.213	0.572	0.000
53	0.085	0.086	0.193	0.000
54	0.083	0.083	0.468	0.000

**Table A11. State X Grade 3 Item-level Word Analysis Ability Proxy Based On All 30 Items**

Item	R-squared values at each step in the sequential hierarchical regression		DIF results	
	Step 1 Ability Proxy	Step 2 Ability Proxy, Disability Status, Interaction	Chi-Sq P-value	Change in R-Square (Effect size)
1	0.347	0.348	0.304	0.001
2	0.317	0.318	0.007	0.001
3	0.401	0.402	0.761	0.001
4	0.251	0.252	0.006	0.001
5	0.200	0.201	0.044	0.001
6	0.458	0.459	0.352	0.001
7	0.410	0.410	0.720	0.000
8	0.138	0.138	0.296	0.000
9	0.152	0.154	0.003	0.002
10	0.505	0.506	0.147	0.001
11	0.433	0.434	0.089	0.001
12	0.354	0.355	0.223	0.001
13	0.391	0.394	0.000	0.003
14	0.485	0.488	0.001	0.003
15	0.207	0.212	0.000	0.005
16	0.426	0.427	0.088	0.001
17	0.323	0.324	0.269	0.001
18	0.462	0.462	0.791	0.000
19	0.204	0.205	0.059	0.001
20	0.182	0.184	0.002	0.002
21	0.321	0.324	0.001	0.003
22	0.259	0.259	0.501	0.000
23	0.358	0.363	0.000	0.005
24	0.246	0.248	0.007	0.002
25	0.237	0.238	0.019	0.001
26	0.352	0.353	0.276	0.001
27	0.167	0.171	0.000	0.004
28	0.141	0.145	0.000	0.004
29	0.398	0.398	0.082	0.000
30	0.412	0.412	0.231	0.000

**Table A12. State X Grade 3 Item-level Word Analysis Ability Proxy Based On First 15 Items**

Item	R-squared values at each step in the sequential hierarchical regression		DIF results	
	Step 1 Ability Proxy	Step 2 Ability Proxy, Disability Status, Interaction	Chi-Sq P-value	Change in R-Square (Effect size)
1	0.379	0.379	0.528	0.000
2	0.378	0.381	0.000	0.003
3	0.444	0.445	0.165	0.001
4	0.317	0.320	0.000	0.003
5	0.266	0.269	0.000	0.003
6	0.488	0.488	0.521	0.000
7	0.435	0.436	0.125	0.001
8	0.206	0.207	0.039	0.001
9	0.211	0.215	0.000	0.004
10	0.307	0.308	0.517	0.001
11	0.448	0.450	0.018	0.002
12	0.358	0.360	0.008	0.002
13	0.429	0.434	0.000	0.005
14	0.494	0.497	0.000	0.003
15	0.249	0.257	0.000	0.008
16	0.342	0.345	0.002	0.003
17	0.217	0.217	0.953	0.000
18	0.358	0.361	0.026	0.003
19	0.116	0.117	0.168	0.001
20	0.087	0.087	0.120	0.000
21	0.180	0.182	0.010	0.002
22	0.149	0.151	0.015	0.002
23	0.206	0.207	0.011	0.001
24	0.136	0.137	0.205	0.001
25	0.137	0.141	0.000	0.004
26	0.220	0.221	0.102	0.001
27	0.080	0.081	0.070	0.001
28	0.065	0.067	0.006	0.002
29	0.262	0.264	0.024	0.002
30	0.277	0.278	0.089	0.001

**Table A13. State Y Grade 3 Item-level Reading Comprehension Ability Proxy Based On All 54 Items**

Item	R-squared values at each step in the sequential hierarchical regression		DIF results	
	Step 1 Ability Proxy	Step 2 Ability Proxy, Disability Status, Interaction	Chi-Sq P-value	Change in R-Square (Effect size)
1	0.419	0.419	0.000	0.000
2	0.531	0.531	0.000	0.000
3	0.401	0.402	0.000	0.001
4	0.230	0.230	0.011	0.000
5	0.547	0.547	0.000	0.000
6	0.142	0.143	0.000	0.001
7	0.337	0.337	0.000	0.000
8	0.079	0.079	0.000	0.000
9	0.100	0.100	0.000	0.000
10	0.311	0.312	0.000	0.001
11	0.353	0.353	0.000	0.000
12	0.390	0.390	0.000	0.000
13	0.282	0.283	0.000	0.001
14	0.245	0.245	0.000	0.000
15	0.067	0.068	0.000	0.001
16	0.352	0.353	0.000	0.001
17	0.338	0.338	0.003	0.000
18	0.019	0.020	0.000	0.001
19	0.203	0.203	0.000	0.000
20	0.332	0.333	0.000	0.001
21	0.414	0.414	0.000	0.000
22	0.286	0.286	0.000	0.000
23	0.305	0.305	0.090	0.000
24	0.398	0.398	0.000	0.000
25	0.548	0.548	0.000	0.000
26	0.385	0.386	0.000	0.001
27	0.349	0.349	0.007	0.000
28	0.279	0.279	0.000	0.000
29	0.369	0.370	0.000	0.001
30	0.355	0.355	0.000	0.000
31	0.429	0.429	0.001	0.000
32	0.365	0.365	0.000	0.000

Item	R-squared values at each step in the sequential hierarchical regression		DIF results	
	Step 1 Ability Proxy	Step 2 Ability Proxy, Disability Status, Interaction	Chi-Sq P-value	Change in R-Square (Effect size)
33	0.155	0.156	0.000	0.001
34	0.487	0.487	0.000	0.000
35	0.352	0.353	0.000	0.001
36	0.496	0.496	0.000	0.000
37	0.398	0.399	0.000	0.001
38	0.241	0.242	0.000	0.001
39	0.281	0.281	0.000	0.000
40	0.362	0.362	0.000	0.000
41	0.345	0.345	0.000	0.000
42	0.435	0.436	0.000	0.001
43	0.427	0.428	0.000	0.001
44	0.224	0.224	0.000	0.000
45	0.343	0.345	0.000	0.002
46	0.236	0.237	0.000	0.001
47	0.288	0.289	0.000	0.001
48	0.312	0.314	0.000	0.002
49	0.482	0.482	0.000	0.000
50	0.213	0.213	0.000	0.000
51	0.191	0.191	0.933	0.000
52	0.426	0.427	0.000	0.001
53	0.162	0.162	0.000	0.000
54	0.205	0.205	0.000	0.000



**Table A14. State Y Grade 3 Item-level Reading Comprehension Ability Proxy Based On First 27 Items**

Item	R-squared values at each step in the sequential hierarchical regression		DIF results	
	Step 1 Ability Proxy	Step 2 Ability Proxy, Disability Status, Interaction	Chi-Sq P-value	Change in R-Square (Effect size)
1	0.484	0.484	0.000	0.000
2	0.610	0.613	0.000	0.003
3	0.476	0.477	0.000	0.001
4	0.267	0.267	0.003	0.000
5	0.618	0.618	0.000	0.000
6	0.188	0.188	0.000	0.000
7	0.377	0.377	0.000	0.000
8	0.110	0.111	0.000	0.001
9	0.136	0.136	0.000	0.000
10	0.332	0.335	0.000	0.003
11	0.400	0.401	0.000	0.001
12	0.422	0.423	0.000	0.001
13	0.334	0.334	0.277	0.001
14	0.261	0.262	0.000	0.000
15	0.082	0.085	0.000	0.003
16	0.349	0.353	0.000	0.004
17	0.371	0.371	0.000	0.000
18	0.026	0.027	0.000	0.001
19	0.228	0.228	0.000	0.000
20	0.354	0.357	0.000	0.003
21	0.434	0.435	0.000	0.001
22	0.309	0.311	0.000	0.002
23	0.345	0.345	0.000	0.000
24	0.416	0.418	0.000	0.002
25	0.535	0.536	0.000	0.001
26	0.378	0.381	0.000	0.003
27	0.369	0.370	0.000	0.001
28	0.237	0.238	0.000	0.001
29	0.295	0.297	0.000	0.002
30	0.293	0.294	0.000	0.001
31	0.351	0.352	0.000	0.001
32	0.309	0.309	0.000	0.000

Item	R-squared values at each step in the sequential hierarchical regression		DIF results	
	Step 1 Ability Proxy	Step 2 Ability Proxy, Disability Status, Interaction	Chi-Sq P-value	Change in R-Square (Effect size)
33	0.102	0.103	0.000	0.001
34	0.385	0.387	0.000	0.002
35	0.271	0.272	0.000	0.001
36	0.404	0.405	0.000	0.001
37	0.292	0.293	0.000	0.001
38	0.161	0.163	0.000	0.002
39	0.203	0.205	0.000	0.002
40	0.266	0.267	0.000	0.001
41	0.250	0.251	0.000	0.001
42	0.335	0.336	0.000	0.001
43	0.316	0.318	0.000	0.002
44	0.146	0.148	0.000	0.002
45	0.233	0.236	0.000	0.003
46	0.152	0.154	0.000	0.002
47	0.197	0.198	0.000	0.001
48	0.219	0.221	0.000	0.002
49	0.371	0.372	0.008	0.001
50	0.147	0.147	0.000	0.000
51	0.139	0.139	0.000	0.000
52	0.317	0.317	0.000	0.000
53	0.108	0.109	0.000	0.001
54	0.144	0.145	0.000	0.001

**Table A15. State Y Grade 3 Item-level Word Analysis Ability Proxy Based On All 30 Items**

Item	R-squared values at each step in the sequential hierarchical regression		DIF results	
	Step 1 Ability Proxy	Step 2 Ability Proxy, Disability Status, Interaction	Chi-Sq P-value	Change in R-Square (Effect size)
1	0.491	0.491	0.371	0.000
2	0.306	0.307	0.000	0.001
3	0.470	0.470	0.000	0.000
4	0.300	0.301	0.000	0.001
5	0.237	0.237	0.000	0.000
6	0.550	0.550	0.000	0.000
7	0.481	0.482	0.000	0.001
8	0.156	0.156	0.000	0.000
9	0.172	0.173	0.000	0.001
10	0.566	0.567	0.000	0.001
11	0.509	0.510	0.000	0.001
12	0.436	0.436	0.000	0.000
13	0.366	0.368	0.000	0.002
14	0.543	0.543	0.000	0.000
15	0.279	0.280	0.000	0.001
16	0.494	0.494	0.000	0.000
17	0.455	0.456	0.000	0.001
18	0.570	0.571	0.000	0.001
19	0.266	0.266	0.171	0.000
20	0.232	0.234	0.000	0.002
21	0.342	0.343	0.000	0.001
22	0.293	0.293	0.780	0.000
23	0.430	0.432	0.000	0.002
24	0.318	0.319	0.000	0.001
25	0.268	0.268	0.117	0.000
26	0.394	0.395	0.000	0.001
27	0.185	0.186	0.000	0.001
28	0.193	0.195	0.000	0.002
29	0.456	0.458	0.000	0.002
30	0.422	0.424	0.000	0.002

**Table A16. State Y Grade 3 Item-level Word Analysis Ability Proxy Based On First 15 Items**

Item	R-squared values at each step in the sequential hierarchical regression		DIF results	
	Step 1 Ability Proxy	Step 2 Ability Proxy, Disability Status, Interaction	Chi-Sq P-value	Change in R-Square (Effect size)
1	0.510	0.511	0.000	0.001
2	0.353	0.356	0.000	0.003
3	0.491	0.491	0.000	0.000
4	0.347	0.350	0.000	0.003
5	0.291	0.292	0.000	0.001
6	0.556	0.559	0.000	0.003
7	0.501	0.502	0.000	0.001
8	0.209	0.213	0.000	0.004
9	0.227	0.231	0.000	0.004
10	0.581	0.582	0.000	0.001
11	0.522	0.523	0.000	0.001
12	0.445	0.448	0.000	0.003
13	0.394	0.400	0.000	0.006
14	0.555	0.556	0.000	0.001
15	0.320	0.324	0.000	0.004
16	0.412	0.415	0.000	0.003
17	0.353	0.354	0.000	0.001
18	0.470	0.478	0.000	0.008
19	0.170	0.171	0.000	0.001
20	0.130	0.131	0.000	0.001
21	0.214	0.215	0.000	0.001
22	0.188	0.189	0.000	0.001
23	0.276	0.277	0.000	0.001
24	0.196	0.197	0.000	0.001
25	0.173	0.176	0.000	0.003
26	0.264	0.266	0.000	0.002
27	0.098	0.099	0.000	0.001
28	0.098	0.099	0.000	0.001
29	0.322	0.324	0.000	0.002
30	0.307	0.310	0.000	0.003