



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

Jamal Abedi, CRESST/University of California, Davis  
Seth Leon and Jenny Kao, CRESST/University of California, Los Angeles

NCME  
April 10, 2007



# Background

- Over 6 million (13%) of all public school students have disabilities
- 44 states reported participation and performance of students with disabilities in their NCLB assessments (Klein, Wiley, & Thurlow, 2006)
- Most students with disabilities participate in regular, not alternate assessments



# Focus on Reading

- Among the content areas, reading is of greater importance.
- Literature shows lower performance on academic achievement tests for students with disabilities (SD).
- These students perform significantly lower than students with no apparent disabilities (Abedi, Leon, & Mirocha, 2003; Thurlow, et. al., 2000; Tindal, et. al, 2000; Ysseldyke et al., 1998)



# SD and Non-SD

- Many factors contribute to performance gap between SDs and non-SDs outside of a student's specific disability (opportunity to learn, testing accommodations, etc)
- We examined the possibility items being potentially biased against students with disabilities
- The results of this study along with the findings of other studies showed differences across SD/non-SD categories
- There are many different forms of detecting item bias



# Test bias

- 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
- Using DIF analyses we can see whether a specific item is biased against a particular group
- Using DDF analyses we can see whether a particular group prefers different distractors



# Data Source

- Two different states
  - State X: small state with average number of SD
  - State Y: large state with average number of SD
- Stanford 9 from 1997-1998 school year
- Item-level information on Grades 3 and 9 Reading Comprehension (RC) and Word Analysis (WA) subscales



# Statistical Design

- A Multi-step Logistic regression approach was used to examine the possibility of both differential item functioning (DIF) and differential distractor functioning (DDF) across the SD and non-SD categories
- SD status and an ability proxy on the subscale being analyzed and the interaction between SD status and ability were used as predictors
- Odds Ratios (OR) of selecting the correct choice (DIF) or the less common distractors (DDF) were computed



# DIF Results-State X

## *Reading Comprehension, Grade 9*

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

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



# DIF Results-State Y

## *Reading Comprehension, Grade 9*

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

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



# DDF Results-State X

## *Reading Comprehension, Grade 9*

Ability Proxy	Total Number of Items	Number of Items Showing DDF		
		Items 1-27	Items 28-54	All Items
Model 1	54	4	6	10
Model 2	54	3	10	13

Of the 10 items that indicated DDF from the second half of the assessment, seven had a significant main effect for the disability status variable. For each of these seven items the odds ratio for students with disabilities was less than 1.0




# DIF Results-State X

## *Word Analysis, Grade 9*

Ability Proxy	Total Number of Items	Number of Items Showing DIF		
		Items 1-27	Items 28-54	All Items
Model 1	30	4	8	12
Model 2	30	6	13	19

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



# DIF Results-State Y

## *Word Analysis, Grade 9*

Ability Proxy	Total Number of Items	Number of Items Showing DIF		
		Items 1-27	Items 28-54	All Items
Model 1	30	0	1	1
Model 2	30	2	10	12

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



# DDF Results-State X

## *Word Analysis, Grade 9*

Ability Proxy	Total Number of Items	Number of Items Showing DDF		
		Items 1-27	Items 28-54	All Items
Model 1	30	4	8	12
Model 2	30	4	7	11

Of the seven items that showed DDF from the second half of the assessment, six had a significant main effect for the disability status variable. For each of these six items the odds ratio for students with disabilities was less than 1.0



# Results

## *Grade 3*

- In Grade 3 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 RC and WA subscales.
- The items that were shown to exhibit DIF for students with disabilities were no more likely to be located in the second half of these assessments than they were in the first half of these assessments.
- Results also indicate that DDF was present for Grade 9 test items, but not for Grade 3 items. Even when controlling for ability using only the items in the first half of the assessments, more Grade 9 items exhibited DDF than Grade 3 items.



# Summary of Findings

- In Grade 9, many items exhibited DIF. These items were more likely to be located in the second half of the test.
- SD consistently under-performed on items located in the second half relative to the first half, as compared with non-SD
- Items showing DDF also were more likely to be located in the second half.
- Suggests SD might be more randomly selecting one of the four response options rather than making an “educated guess.”



# Limitations

- This study did not differentiate between different categories of disabilities
- Did not have access to information on testing accommodations
- Actual test items were not available for examination
- We could not identify other test characteristics



# More Research Needed

- Why there are differences between Grades 3 and 9
- What factors contribute to the diminishing performance for SD as the test progresses
- Studies differentiating between different types of disabilities



# Conclusions

- Findings of this study provide evidence that in addition to test content other factors may contribute to the performance gap between students with disabilities and their non-disabled peers
- Controlling for these factors that are not related to content being assessed may help test developers provide more accessible and more valid assessments for students with disabilities



For more information contact:

Jamal Abedi

[jabedi@ucdavis.edu](mailto:jabedi@ucdavis.edu)

(530) 754-9150