



# Research on Making Large Scale Assessments More Accessible for Students with Disabilities

Institute of Education Sciences Research Conference  
June 7-8, 2007

Cara Cahalan Laitusis and Linda Cook  
Educational Testing Service



# Designing Accessible Reading Assessments (DARA)

- Educational Testing Service (ETS)
- Focus on students with learning disabilities
- Focus on component approach to assessing reading skills. Primary areas are:
  - Word Recognition
  - Reading Fluency
  - Vocabulary Knowledge
  - Comprehension



# Today's Presentation

- Who is reading aloud NCLB reading assessments and why?
- Review of research examining the impact of reading aloud modifications on the validity of test scores.
- Summary of findings and implications for developing an accessible reading assessment



# Who is Reading the Reading Test?

States are divided into four positions

- California and New Jersey: Reading involves decoding (visual or tactile) so read aloud is not allowed
- Wisconsin: Read aloud results in test of another construct (listening comprehension) and test was not designed for this purpose so read aloud is not allowed
- Kentucky and Delaware: Reading is comprehension of text (visual, tactile, or audio) so read aloud is allowed
- Texas and Maine: Word level decoding is not the primary construct so some read aloud is permitted (e.g., reading test questions but not passages)



# Prevalence of Read Aloud

- Many students use read aloud on tests of reading even when it is not allowed under their state's policy
  - In 2006 nearly 5000 4<sup>th</sup> graders in California used Read Aloud on the STAR English Language Arts assessment even though the accommodation is “not allowed” and invalidates the student's score for NCLB reporting.
- When read aloud is permitted the use of read aloud can be very large (up to 70% of students with disabilities) leading some states to change their policy or list specific caveats in 2006.



# Read Aloud Policy Changes in 2006

- Texas changed policy to permit read aloud (e.g., reading aloud proper nouns, test questions, and answer choices but not passages) on some reading tests for dyslexic students only
- Massachusetts changed policy to allow screen reading software, but specifies the accommodation for students with disabilities who are unable to read, not those performing below grade level
- Missouri changed policy to discontinue allowing read aloud due to large number of students using the accommodation (70% of students with disabilities)



# Research Questions

- Do students with reading-based learning disabilities receive differential performance gains from read aloud or would all poor readers benefit from having the test read aloud?
- Do tests and test items taken with and without read aloud perform the same psychometrically?
  - same factor structure
  - no evidence of differential item performance
  - equivalent predictions of reading comprehension



# DARA Research on Read Aloud

- Experimentally Designed Study
  - Differential Boost from Read Aloud
  - Predictive Validity of Scores from Read Aloud
- Psychometric Studies
  - Differential Item Functioning
  - Differential Distractor Analysis
  - Factor Analysis
- Think Aloud Studies



# Differential Boost Description

- Framework for studying impact of changes in testing conditions
- Change in testing condition may be viewed as an accommodation if students with disabilities receive a significant increase in scores compared to students without disabilities
- Most researchers refer to differential boost as an interaction between group and change in testing condition



# Data Collected

- 2 Reading Comprehension Tests
  - GMRT Forms S & T
  - Audio Form (also had extra time and answering in test booklet)
  - Standard Form (extra time and answering in test booklet)
- 2 Fluency Measures
  - WJ Reading Fluency and TOSWRF
- 2 Decoding Measures (4<sup>th</sup> grade only)
  - WJ Letter Word ID and WJ Word Attack
- Student Survey
- Teacher Survey



# Sample

- 1181 4<sup>th</sup> grade students
  - 527 students with reading based learning disabilities (RLD)
  - 654 students without a disability (NLD)
- 855 8<sup>th</sup> Graders
  - 394 RLD
  - 461 NLD



# Differential Boost Data Collection Design

Group	Session 1		Session 2	
	Form	Accommodation/ Modification	Form	Accommodation/ Modification
1	S	Standard	T	Audio
2	S	Audio	T	Standard
3	T	Standard	S	Audio
4	T	Audio	S	Standard



## Grade 4

	RLD (n=527)		NLD (n=654)	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Standard	457	32	497	38
Audio	477	30	502	33
<b>Boost</b>	<b>20</b>	<b>29</b>	<b>5</b>	<b>24</b>
Fluency	473	21	500	25



# Grade 8

	RLD (n=376)			NLD (n=471)	
	<i>M</i>	<i>SD</i>		<i>M</i>	<i>SD</i>
Standard	511	28		553	33
Audio	521	27		555	31
<b>Boost</b>	<b>10</b>	<b>23</b>		<b>2</b>	<b>21</b>
Fluency	514	34		560	42



# Differential Boost Analyses

## RM-ANOVA

- Dependent Variable: Boost
- Independent Variables:
  - Experimental Group (form/accommodation order)
  - LD Status (RLD or NLD)

## RM-ANOVA

- Controlling for ceiling effect by removing top performers
  - ( $\geq 45$  of 48 items correct on standard)

## RM-ANCOVA

- Covariate: WJ Reading Fluency



# Results of the Differential Boost Analyses

- Students with reading-based learning disabilities had a significantly larger boost from audio (read aloud) accommodations than students without disabilities
  - Controlling for other factors (e.g., reading fluency, decoding, or ceiling effects) does not change these findings



# Predictive Validity Analyses

Regression analyses examining which test scores capture the most variance in teachers rating of reading comprehension

- Tested 4 models:
  - Standard
  - Standard + Fluency
  - Audio
  - Audio + Fluency
- Examined which assessment model should be used to assess reading comprehension by grade and disability group
- Several limitations to these analyses
  - Ratings were done in October
  - Definition of reading comprehension may have varied between teachers
- Even with these limitations this study is the first study to examine the predictive validity of audio (read aloud) test scores



# Results from Regression Analyses

- Standard score alone is a good measure of reading comprehension (as defined by teachers) for NLD students and an adequate measure for RLD students

$\beta = .607$  for grade 4 NLD and  $.525$  for grade 8 NLD

$\beta = .460$  for grade 4 NLD and  $.405$  for grade 8 NLD

- Audio score does not appear to predict teachers ratings of reading comprehension better (or as well) than the standard score for RLD students

Audio score with fluency (and standard with fluency) captured more variance in teachers ratings than standard score alone for Grade 4 RLD



# Purpose of Differential Item Functioning and Factor Analyses

- Determine if the test is measuring the same underlying construct (s) for the following comparisons:
  - Students without disabilities who take the test with and without a read aloud test change
  - Students with disabilities who take the test with and without a read aloud test change



# Differential Item Functioning Analyses

- Comparisons of item responses for four groups
  - Reading based learning disability (RLD), no accommodation (Standard)
  - Reading based learning disability (RLD), audio modification (Audio)
  - No disability (NLD), no accommodation (Standard)
  - No disability (NLD), audio modification (Audio)



# Summary Statistics for Grade 4 Form S DIF and Factor Analysis Samples

Comparison	Reference Group	N	Mean	SD	Focal Group	N	Mean	SD
1	RLD Standard	258	19.18	9.05	RLD Audio	269	24.36	8.81
2	NLD Standard	326	30.08	9.68	NLD Audio	328	32.44	8.81



# Differential Item Functioning (DIF)

- Test takers matched on proficiency level
  - Used total test score as matching criterion
- Reference Groups
  - Students with and without disabilities who took test under standard conditions
- Focal Groups
  - Students with and without disabilities who took test with read aloud test change

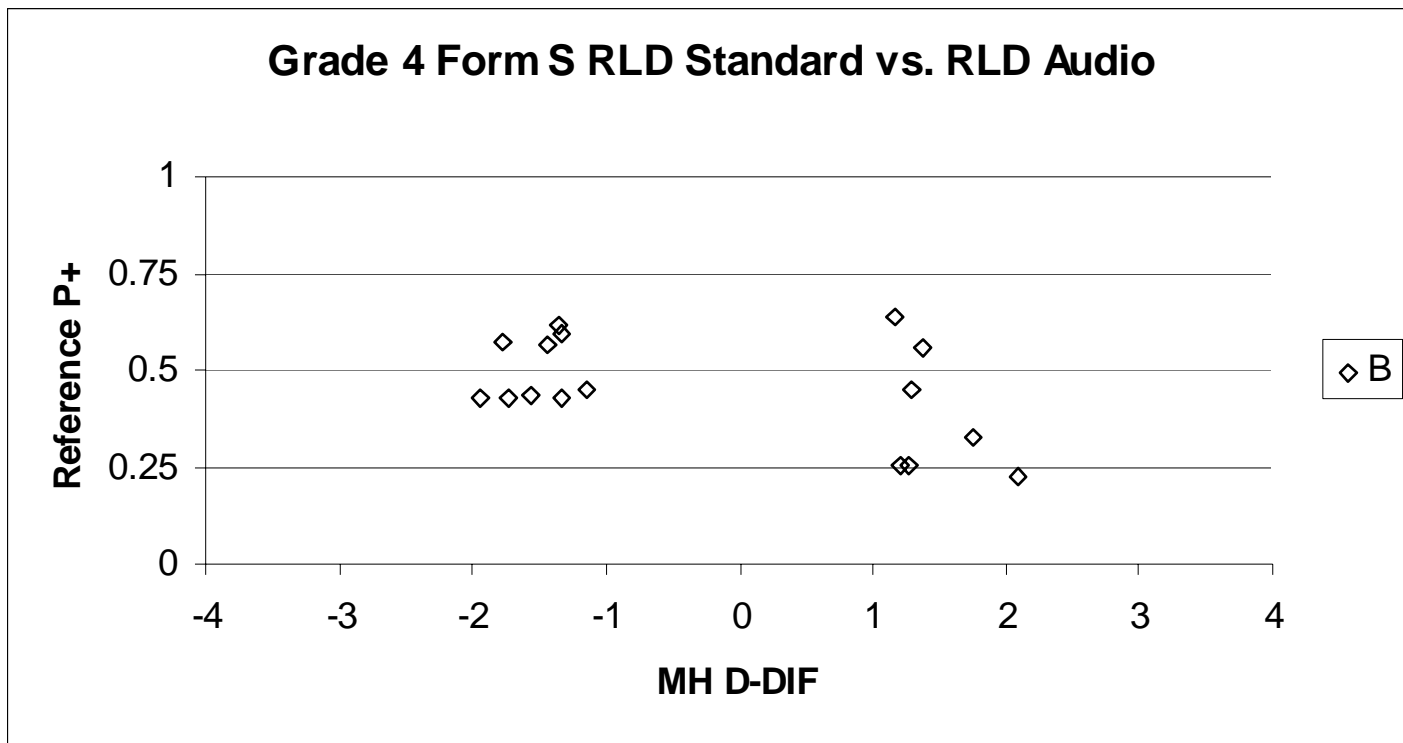


# Differential Item Functioning (DIF) Analyses

- Used Mantel-Haenszel procedure with total score as criterion
- Mantel-Haenszel categorization
  - A—negligible DIF
  - B—slight to moderate DIF
  - C—moderate to large DIF
- Direction of DIF Flags
  - Negative favors reference group
  - Positive favors focal group

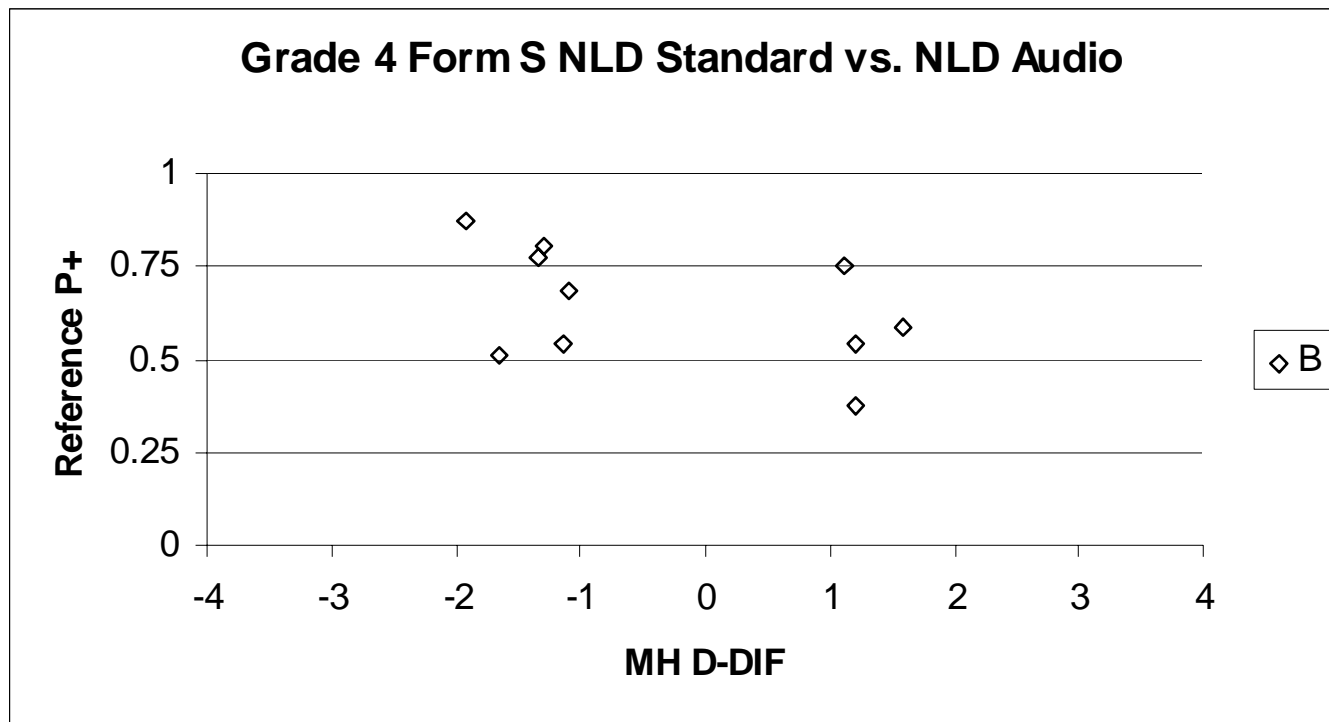


# Comparison 1: RLD Standard (reference group) vs. RLD Audio (focal group)





# Comparison 2: NLD Standard (reference group) vs. NLD Audio (focal group)





# Factor Analysis of GMRT Data

- Analyzed Form S (4<sup>th</sup> and 8<sup>th</sup> grade levels) and Form T (4<sup>th</sup> and 8<sup>th</sup> grade levels)
- Investigated the impact of a read aloud test change for students with and without disabilities who took the test with and without the test change
- Does the GMRT measure the same underlying construct (s) for the study groups?



# Factor Analysis of GMRT Data

- Single group exploratory factor analysis of four groups
  - RLD who took the test under standard conditions
  - RLD who took the test with an audio test change
  - NLD who took the test under standard conditions
  - NLD who took the test with an audio test change
- Confirmatory factor analysis
  - Single-group to establish single factor
  - Multi-group to establish base line model
  - Multi-group comparisons to establish measurement invariance
    - Comparison 1—RLD standard vs. RLD read aloud
    - Comparison 2—NLD standard vs. NLD read aloud



# Summary of Multi-group CFA for RLD Standard vs. RLD Audio

Summary of CFA Testing Equality of Factor Loadings, Variances, and Residuals for RLD Standard vs. RLD Audio							
Model	DF	Normal Theory Chisq.	Normal Theory Chisq./DF	RMSEA	GFI	CFI	Change in CFI
Multi-Group CFA (3a)	40	58.325	1.458	.042	.973	0.990	--
Multi-Group CFA (3b)	47	66.916	1.424	.040	.970	0.989	.001
Multi-Group CFA (3c)	48	67.044	1.397	.039	.970	0.989	.000
Multi-Group CFA (3d)	56	72.604	1.297	.034	.967	0.991	-.002



## Summary

- Purpose of the studies was to determine if the test measures the same underlying construct (s) when taken with and without a read aloud test change
- Results of DIF and factor analysis studies indicate that test measures the same underlying construct for the RLD group, or for the NLD group when taken with and without a read aloud test change



# Additional Research Studies

- Pitoniak, M., Cook, L., Cline, F., & Cahalan-Laitusis, C. *Using differential Item Functioning to Investigate the Impact of Accommodations on the Scores of Students with Disabilities on English-Language Arts Assessments.*
- Cook, L. Eignor, D., Sawaki, Y., Steinberg, J., & Cline, F. *Using factor analysis to investigate the impact of accommodations on the scores of students with disabilities on English-language arts assessments.*
- Bolt, S. E., & Diao, Q. *Reading-aloud a reading test: Examining reading sub-skill performance.*
- Koretz, D. *The assessment of students with disabilities in Kentucky.*
- Koretz, D., & Hamilton, L. *Assessing students with disabilities in Kentucky: the effects of accommodations, format, and subject.*



# Research Answers

- Students with reading-based learning disabilities DO receive differential performance gains from read aloud.
  - These gains can not be completely explained by poor fluency ability or ceiling effects
- Reading tests and test items taken with and without read aloud:
  - Have the same factor structure
  - Have no evidence of differential item performance
  - But do not predict reading comprehension in the same way for students with and without disabilities



# Implications for Accessible Reading Assessments

- Based on regression findings assessments should test access skills (fluency and decoding) separately from comprehension for RLD students
- Based on DIF and factor analyses, a reading test given with an audio test change appears to measure the same constructs as a reading test given under standard conditions
- Audio score alone may be an easier but not a better measure of 'reading comprehension' (compared to standard score) for RLD students



# Questions?

Cara Cahalan Laitusis, [ccahalan@ets.org](mailto:ccahalan@ets.org)

Linda Cook, [lcook@ets.org](mailto:lcook@ets.org)

or

visit [www.ets.org/dara](http://www.ets.org/dara)