



Designing Accessible Reading Assessments (DARA) A Summary of Research

Teresa King
Educational Testing Service



DARA Research Questions

1. Do students with reading-based learning disabilities receive differential performance gains from read aloud when compared to students without disabilities?
2. Do reading comprehension tests and test items taken with and without a read aloud measure similar constructs?



DARA Research Studies

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



Summary of DARA Analyses

Analyses	Test/Sample			
	GMRT Grade 4	GMRT Grade 8	ELA Grade 4	ELA Grade 8
Differential Boost/Predictive Validity	X	X		
Differential Item Functioning	X	X	X	X
Factor Analysis	X	X	X	X
Distractor Analysis			X	X
Cognitive Labs	X	X		



Differential Boost and Predictive Validity Research Questions

1. Does a read-aloud accommodation provide a differential boost for students with learning disabilities?
2. How well do test scores predict teacher's ratings of a student's reading comprehension level?
3. Are teachers able to predict which students will benefit from a read-aloud accommodation?



Differential Boost Methods

Primary Measure

- 2 GMRT Reading Comprehension Tests (Form S and Form T)
 - Extra time (Standard)
 - Extra time with Read-Aloud via CD (Audio)

Additional Measures

- 2 Fluency Measures
- 2 Decoding Measures (4th grade only)
- Student Survey
- Teacher Survey

Sample

- 1181 4th grade students
 - 527 students with reading-based learning disabilities (RLD)
 - 654 students without a disability (NLD)
- 855 8th 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 Differential Boost Analysis

Differential Boost Study: Grade 4 Summary Statistics for GMRT						
	Non-LD			RLD		
Condition	N	Mean	SD	N	Mean	SD
Audio	654	502	33	527	477	30
Standard	654	497	38	527	457	32
Boost	654	5	24	527	20	29



Results of the Differential Boost Study

- Students with reading disabilities had a significantly larger boost from an audio (read-aloud) accommodation than students without disabilities (findings consistent with Fletcher et al. 2005 and Crawford & Tindal, 2004).
 - Controlling for other factors (e.g., reading fluency, decoding, or ceiling effects) does not change these findings
 - Significant interaction between score gain (boost) and disability status was observed
- Answer to first question: Students with reading-based learning disabilities do exhibit differential performance gains when they take a reading test with a read-aloud test change



Predictive Validity of Scores

- Regression analyses were conducted to examine which test scores captured the most variance in teachers ratings of reading comprehension by grade and disability group.
- Tested 4 models:
 - Standard
 - Standard + Fluency
 - Audio
 - Audio + Fluency
- Answer to second question:
 - Tests taken with read-aloud do not predict teachers ratings of reading comprehension as well as tests taken under standard conditions.
 - However combining read-aloud scores with reading fluency scores results in equal (or better) predictions of teacher ratings than tests taken under standard conditions.



Differential Item Functioning (DIF)

- Purpose: to examine whether or not the assessment measured the same construct(s) for the groups in our study
- 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



Differential Item Functioning

- 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 and Factor Analysis of GMRT Data

- 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)



Factor Analyses

Purpose: to examine whether or not the assessments (GMRT and State ELA) measure the same construct(s) for the groups in our study

- Exploratory analyses (separately in each group)
 - How many factors?
- Confirmatory (multi-group)
 - Establish base-line model
 - Confirm number of factors needed to describe data across all groups



Summary of GMRT Analyses

- Results of the DIF study indicated very few C level DIF items (1-2% of items). More items (10-15%) were flagged for B level DIF but the direction of the DIF was equally distributed between the reference and focal groups.
 - Comparisons focusing on mode of administration showed little DIF
 - Comparisons focusing on disability status showed little DIF
 - Comparisons that crossed mode of administration and disability status showed substantial DIF
- Results of the factor analysis studies the test is measuring the same single factor for all subgroups.



Summary of State ELA Analyses

- Results of DIF analyses indicated the level of DIF was less than the DIF found in the GMRT analyses but did not include a non-disabled read-aloud comparison group.
- Results indicated same single factor structure for all of the subgroups studied.



Read Aloud Implications

- 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



Cognitive Labs

Research questions

1. What are the differences in the errors on reading comprehension for students with and without disabilities?
2. Do students with reading-based disabilities differ in their explanations to reading comprehension questions compare to students without disabilities?
3. What kind of information does a cognitive lab provide for future test development reading assessment for students with and without reading-based learning disabilities?
4. Is this type of information gathering and data quality worthwhile to conduct in future large scale studies?



Cognitive Labs Methods

- 49 students participated from 5 NJ schools
 - 34 4th grade students (16 RLD, 18 NLD)
 - 15 8th grade students (7 RLD, 8NLD)
- A think aloud protocol was followed for individual test administration
- Students completed:
 1. Comprehension section of the GMRT
 2. Measures of oral reading fluency
 3. Student survey



Cognitive Labs Findings

4th Grade

RLD correct = 55%

NLD correct = 84%

8th Grade

RLD correct = 35%

NLD correct = 71%

- Findings were grouped into two categories:
 - Text-Based
 - Test-Based
- Examples of differences for RLD students:
 - There was a pattern of RLD students getting test questions correct using the incorrect rationale
 - RLD students were much more likely to use outside knowledge to help support responses (that did not necessarily apply)
 - RLD students were less likely to refer back to the passage to select an answer.



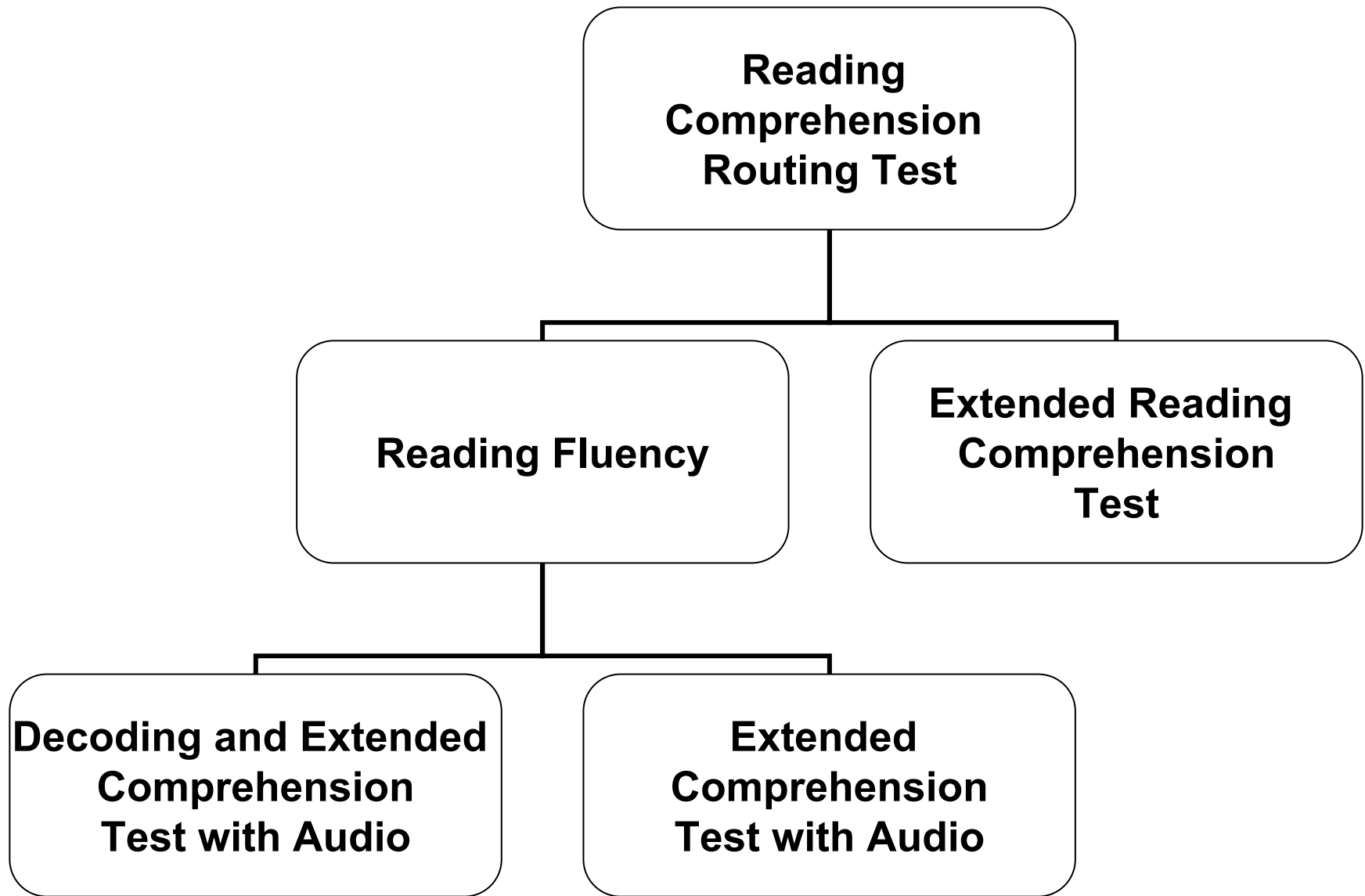
Cognitive Lab Implications

- This methodology is helpful to gain insight on the specific errors student make given a component approach to measuring reading
 - When there was an error, it was easy to identify the problem.
- This procedure will be useful to test new item types and specific subject areas on target populations (LD, VI, blind, hearing impaired, etc.)
- This is a good means to examine non content related aspects of text and testing such as motivation, interest level, and prior knowledge.



Next Steps

- Can we design a multi-stage reading assessment that assesses different components of reading (decoding, fluency, comprehension of text) in isolation (and simultaneously) and report scores on the same scale?
 - Test 1: Audio comprehension, decoding, oral reading fluency
 - Test 2: Comprehension without read aloud
- Conduct simulation studies using real data to examine test reliability and comparability on multi-stage reading assessment.





Questions?