



Designing Accessible Reading Assessments (DARA)

Cara Cahalan-Laitusis
Educational Testing Service



- Overview of DARA Project
- Research Results . . . so far
- Future Research Plans



DARA Project

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



Consultants/Subcontractors

- Jerry Tindal, University of Oregon
- Laurie Cutting, Kennedy Krieger Institute
- Council for Exceptional Children
- Westat



Primary Questions for Year 2

- Can comprehension be assessed in audio format if word recognition and fluency are assessed separately?
 - Are listening comprehension and reading comprehension similar constructs (highly correlated) in proficient readers?
 - Do students with reading-based learning disabilities receive differential performance gains from read aloud?
 - Do tests and test items taken with and without read aloud perform the same psychometrically (same factor structure, equal item difficulty levels)?



Year 2 Research

- Differential Boost from Read Aloud on Reading Test
- Differential Item Function on ELA test
- Factor Analysis of ELA test



Is Read Aloud Allowed?

- States are not in agreement on read aloud on reading tests
- Many students use read aloud even when “not allowed” by the state
 - In California nearly 4,000 fourth graders took the STAR English Language Arts assessment in 2004 with the test read aloud
- Schools are torn between complying with state policy to not allow read aloud and federal regulations to allow all accommodations used in the classroom



Prior Research

- No Differential Boost
 - Koscioczek & Ysseldyke (2000)- Small sample size (n=31)
 - Meloy, Deville, and Frisbie (2002) – Between subjects design (n=260, 76% non-disabled, randomly assigned to audio or standard)
 - McKevitt & Elliott (2003)-Small sample size (n=39)
- Differential Boost
 - Crawford and Tindal (2004)-(n=338, 78% non-disabled)
 - Fletcher, et. al (2006)-Between subjects design (randomly assigned to audio or standard). Sample included 91 Dyslexic (poor decoder) and 91 average decoders



Differential Item Functioning

- Read Aloud on ELA/Reading tests results in more items with DIF than other testing accommodations or read aloud on Math tests:
 - Bielinski, Thurlow, Ysseldyke, Freidebach & Freidebach (2001)
 - Bolt & Bielinski (2003)
 - Cahalan-Laitusis, Cook, & Aicher (2003)



Study 1: Differential Boost from Read Aloud (Non-disabled vs. RLD)

1. Differential Boost from read aloud
2. How well test scores (standard, audio, and fluency) predict variance in teacher ratings of reading comprehension
3. Teachers' ability to predict which students would benefit from read aloud



Data Collected

- 2 Reading Comprehension Tests
 - Extra Time
 - Extra Time with Read Aloud via CD
- 2 Fluency Measures
- 2 Decoding Measures (4th grade only)
- Student Survey
- Teacher Survey



Sample

- 1170 4th Graders
 - 522 Students with RLD
 - 648 Students without a disability

- 855 8th Graders
 - 394 Students with RLD
 - 461 Students without a disability



Design

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

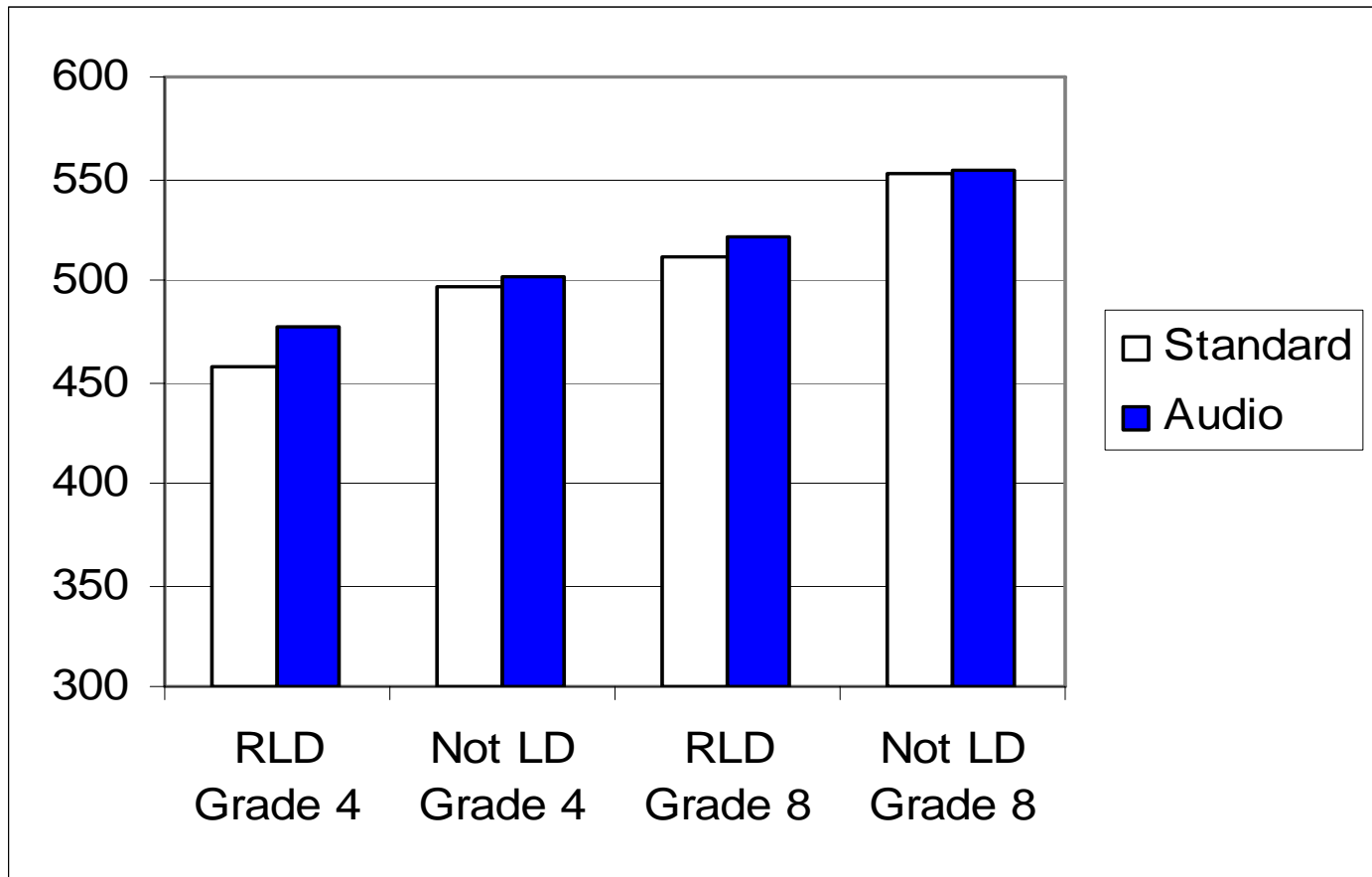


Differential Boost

- RM-ANOVA
- Dependent Variables:
 - GMRT “Standard”
 - GMRT Audio
- Independent Variables:
 - Disability Status (RLD vs. NLD)
 - Form/Order (STSA, STAS, TSSA, TSAS)
- Covariate: WJ-III Reading Fluency



Scores by RLD and Grade





Scores by Grade/RLD

	Grade 4			
	RLD (n=527)		NLD (n=654)	
	<u>M</u>	<u>SD</u>	<u>M</u>	<u>SD</u>
Standard	456.6	32.0	496.9	37.5
Audio	476.7	30.0	501.9	32.5
Boost	20.1	29.0	5.0	23.7
Fluency	473.3	20.7	500.4	24.6

	Grade 8			
	RLD (n=376)		NLD (n=471)	
	<u>M</u>	<u>SD</u>	<u>M</u>	<u>SD</u>
Standard	510.8	27.6	552.8	32.9
Audio	520.6	27.3	554.7	30.5
Boost	9.8	22.9	2.0	20.8
Fluency	513.6	33.6	560.0	41.7



RM ANOVA for Grade 4

Repeated Measures Analysis of Variance for Grade 4

Source	df	F	p
Within subjects			
Boost	1	265.81***	.000
Boost x Reading LD	1	96.46***	.000
Boost x Form/Order	3	0.62	.602
Boost x Reading LD x Form/Order	3	1.35	.258
Error(Boost)	1,173	(342.85)	

Note. Value enclosed in parentheses represent mean square errors.

* $p < .05$. ** $p < .01$, *** $p < .001$.



RM ANOVA for Grade 4 with Fluency Covariate

Table 8. Repeated Measures Analysis of Variance for Grade 4 with Fluency

Source	df	F	p
Within subjects			
Boost	1	71.43***	.000
Fluency (Covariate)	1	58.87***	.000
Boost x Reading LD	1	22.50***	.000
Boost x Form/Order	3	0.91	.438
Boost x Reading LD x Form/Order	3	1.50	.213
Error(Boost)	1,171	(323.03)	

Note. Value enclosed in parentheses represent mean square errors, * $p < .05$.

** $p < .01$, *** $p < .001$



Differential Boost Findings

- Differential Boost at both 4th and 8th grades (i.e., students with LD had significantly greater score gains from read aloud than non-LD students)
- When reading fluency ability is controlled for a Differential Boost is still found at both grades



Regression Findings

- Audio score does not significantly predict variance in Teacher Ratings of Reading Comprehension (beyond standard and fluency) for Grade 8 RLD
- Audio score adds to prediction of reading comprehension (beyond standard and fluency scores) for three groups (NLD grade 4, NLD grade 8, and RLD grade 4), but incremental change is small



Findings from Teacher Predictions

- On average teachers were able to predict score gain from audio at grade 4 but not grade 8
- At the individual level teachers accurately predicted if a student would benefit from the audio version about 35% of the time and were completely wrong about 5% of the time



Study 2: Differential Item Functioning

- Issues investigated:
 - Do 2 different DIF detection methods (Mantel-Haenszel and Logistic Regression) yield the same results?
 - Are the results interpretable in terms of a priori or a posteriori evaluation of item content?
 - When the read-aloud modification is used, do the items function differentially?



Reference-Focal Comparisons

- Nondisabled Standard (NDS) vs. LD Standard (LDS)
- NDS vs. LD Read Aloud
- NDS vs. LD Allowable Accommodation
- LDS vs. LD Read Aloud
- LDS vs. LD Allowable Accommodation



DIF Results

- Very few items resulted in DIF, but Read Aloud items showed the most DIF
- A priori theories of DIF were not accurate predictors
- The LR and MH procedures resulted in some different items being flagged for DIF



Study 3: Factor Analysis

- Does the assessment measure the same number of underlying construct (s)
 - for examinees with a learning disability who take the test with or without an accommodation as it does for the non-disabled population?
 - for examinees with a learning disability who take the test with an accommodation as it does for the examinees with a learning disability who take the test without an accommodation



Factor Analysis design/method

- A series of item level exploratory and confirmatory factor analyses were carried out.
 - The number of factors for the ELA assessment for each group was determined separately
 - The equivalence of the number of factors across groups was tested in a multi-group analysis.



Factor Analysis Results

- Indicate one factor needed to account for data
- Support use of total test score as matching criterion for DIF study
- Support aggregation of scores for AYP purposes
- Provide some evidence of validity of test for students with disabilities
- Further analyses is required to determine if single factors are the same factor for all groups of interest



Are listening/reading similar constructs for RLD and NLD students?

- For NLD students reading and listening are highly correlated at grade 4 (.78) and grade 8 (.79), but less correlated for RLD (.56 and .65)
- Including audio sub-scores will add more predictive validity at grade 4 than grade 8, particularly for RLD grade 4.



Is there a differential boost from read aloud?

- Read aloud does offer a differential performance boost at both grade 4 and grade 8, even when reading fluency is controlled for.
- The boost is larger at grade 4 than grade 8 indicating that the benefit of read aloud may plateau.



Are test items with and without read aloud psychometrically comparable?

- Read Aloud does appear to change item difficulty level for some ELA items
- More research on factor structure and item difficulty is needed



Year 2 Ongoing Research

- Additional Analyses of Differential Boost Data
 - Examine which variables (standard score, fluency, teacher ratings, student preference) best predict boost from audio
- Expand DIF Research
 - Compare use of SIBTEST to MH and LR
 - Differential Boost data file to match both samples on accommodated or standard test score



Year 3 Research

- Think Aloud studies with LD and non-LD students to examine how students approach
 - items shown to have DIF
 - new item types designed to assess fluency and word recognition in a large scale assessment
 - Families of items with slight variations (e.g., operational item and universally designed item)