



Multistage Testing

Cara Cahalan Laituis
Educational Testing Service



Today's Presentation

- Issues with current state reading assessments
- Review of the DARA Project
- Review of multi-stage testing and why DARA project thinks this is a good option for an accessible reading assessment
- DARA projects proposed design for multi-stage test format
- Questions we will focus on for Year 3.



Issues in K12 Reading Assessments for Students with Disabilities

- NCLB assessments (based on grade level achievement standards) are too difficult to provide reliable scores for some students with disabilities
 - About 20% of 4th and 8th graders with learning disabilities in the DARA research studies scored at the chance level on grade level reading tests
 - the ELA portion of a large state reading assessment
 - the reading comprehension portion of a norm referenced test



Issues in K12 Reading Assessments for Students with Disabilities

- Testing with items that are too challenging wastes instructional time for both teachers and students
 - 20-30% of the operational ELA test items (from the DARA research studies) have an average percent correct at chance level (or lower) for students with learning disabilities.
- Many test questions are too difficult which can result in decreased motivation, frustration, and lower self esteem in students.



Operational ELA Assessment

Grade	Disability	Format	Percent Below Chance	
			Test Takers	Items
4	LD	Audio	17.4	21.3
		Other	23.3	30.7
		Standard	21.2	21.3
	None	Standard	2.2	1.3
8	LD	Audio	20.0	32.0
		Other	19.7	32.0
		Standard	16.6	24.0
	None	Standard	1.5	0.0



Issues in K12 Reading Assessments for Students with Disabilities

- Some testing accommodations result in test scores not being counted because the validity of test scores is questioned
 - In 2006 nearly 5000 4th graders in California did not have their AYP scores counted because they had the ELA test read aloud.



Designing Accessible Reading Assessments (DARA) Project

- 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



Advantages of Component Approach

- Allows for the use of read aloud on comprehension and vocabulary constructs
- Provides more diagnostic information for teachers, parents, and school administrators
- Allows student to demonstrate proficiency on some standards even if they are not proficient at grade level on all the standards (e.g., proficient in comprehension but not decoding)



Disadvantages of Component Approach

- Does not measure the integration of each of the components.
- Takes additional time
- Requires individual administration or computer administration
- Requires additional technology (e.g., CD of test or computer administration)



Multi-Stage Testing Solution

- Provides component approach only after student responds poorly to integrated (standard) approach
- Reduces the test time required by lengthy one-on-one administrations required by read aloud via a human reader
- Only requires additional testing time (to assess reading fluency and decoding) for a small number of students
- Allows reuse of test items from standard assessment



Advantage of Multistage (compared to P&P)

- Allows for more efficient and precise measurement across the proficiency scale
- Reduces testing and score reporting time
- Provides equal (or higher) concurrent and predictive validity

From Hendrickson, A. (2007). *An NCME Instructional Module on Multistage Testing*. Educational Measurement: Issues and Practice, Summer 2007, pp 44-52.



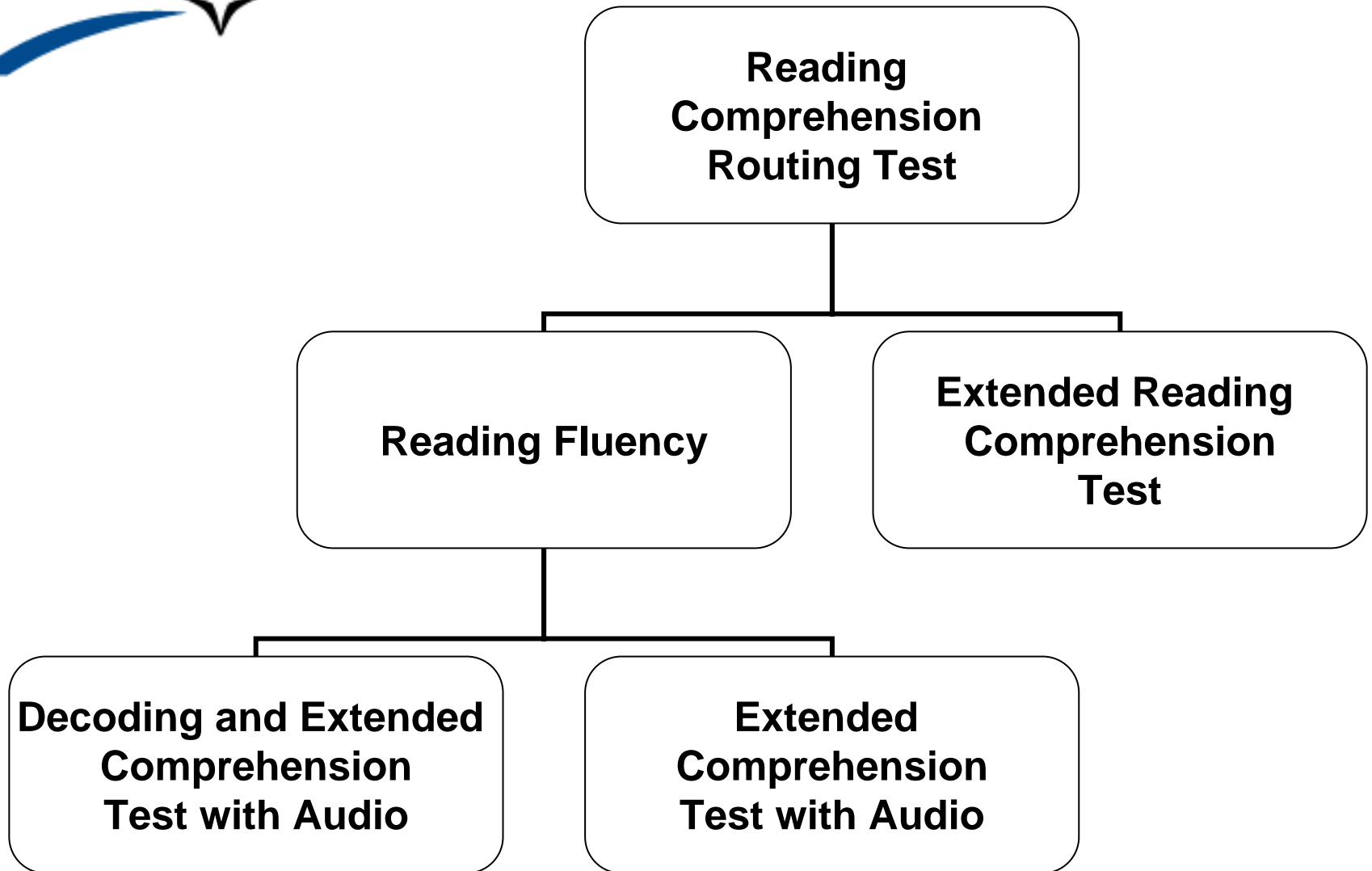
Multi-stage Testing: Purpose

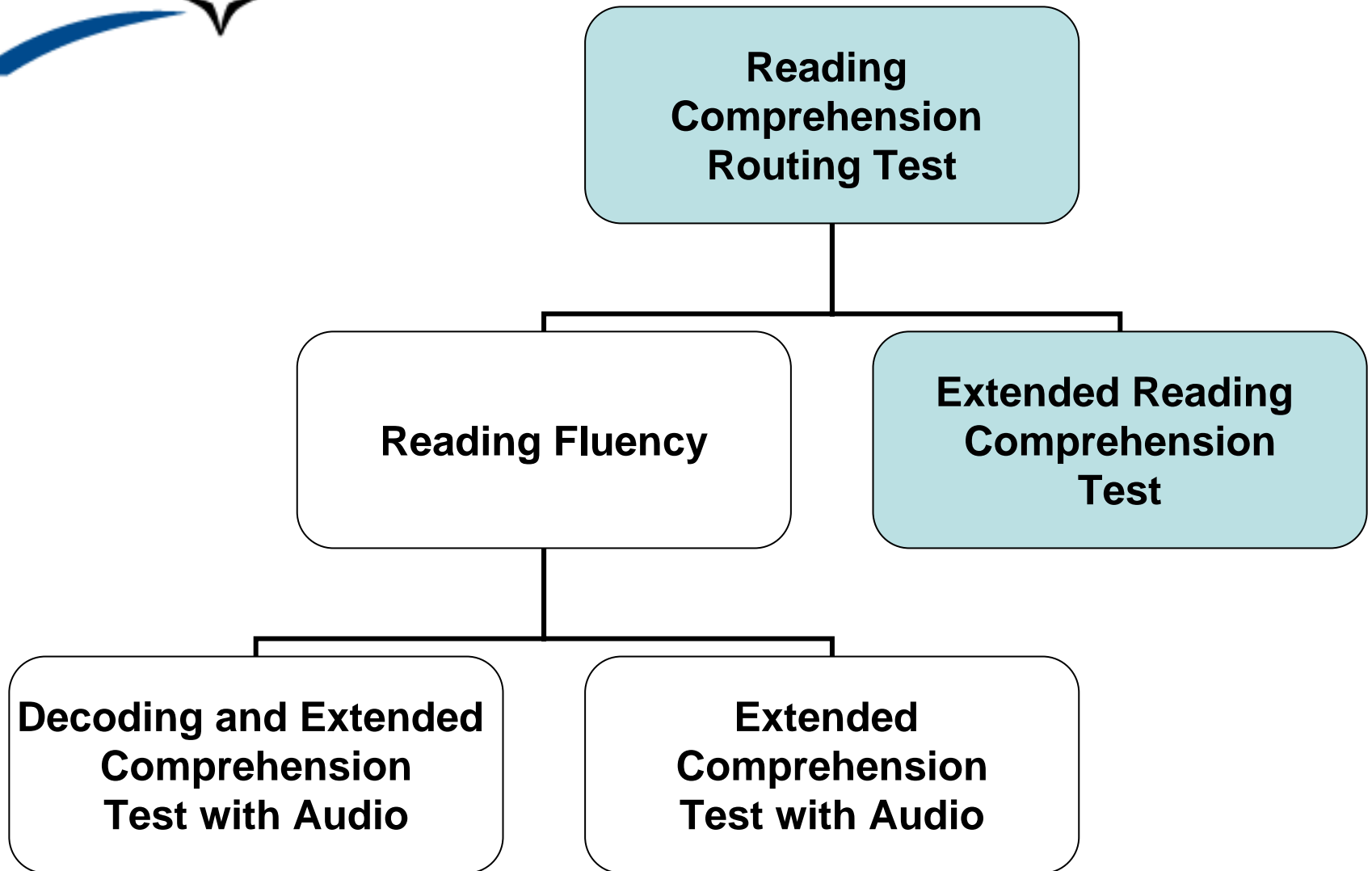
- To investigate ways to provide meaningful scores to students with learning disabilities and to improve the reliability and validity of scores on grade level assessments for students with disabilities through the use of two-stage testing.
 - Isolation of comprehension and decoding/fluency into separate test items
 - Allowing students to demonstrate comprehension via audio (read aloud) format
 - Tailored item/testlet difficulty to students ability without the challenges of a computer adaptive test (CAT)

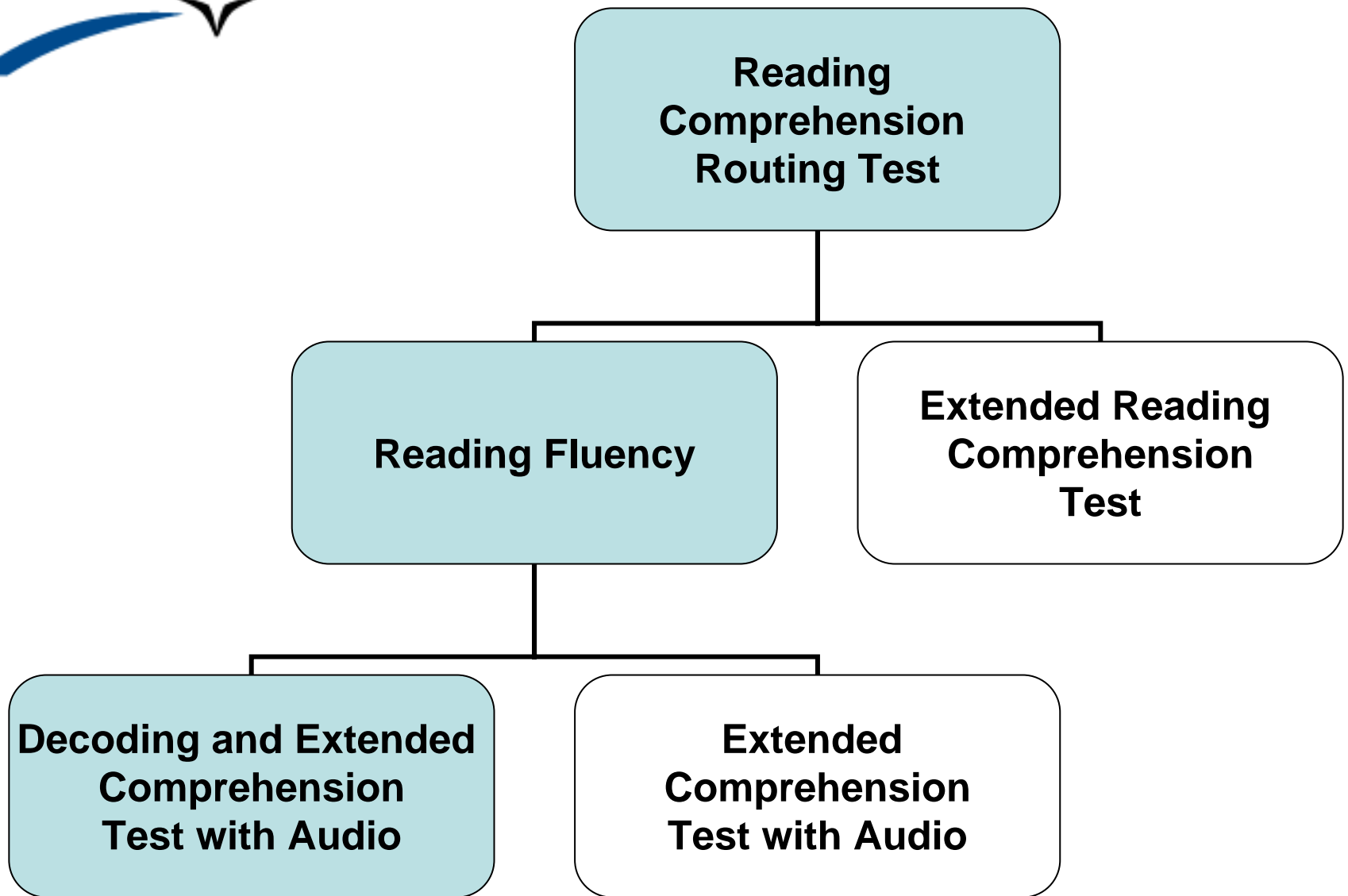


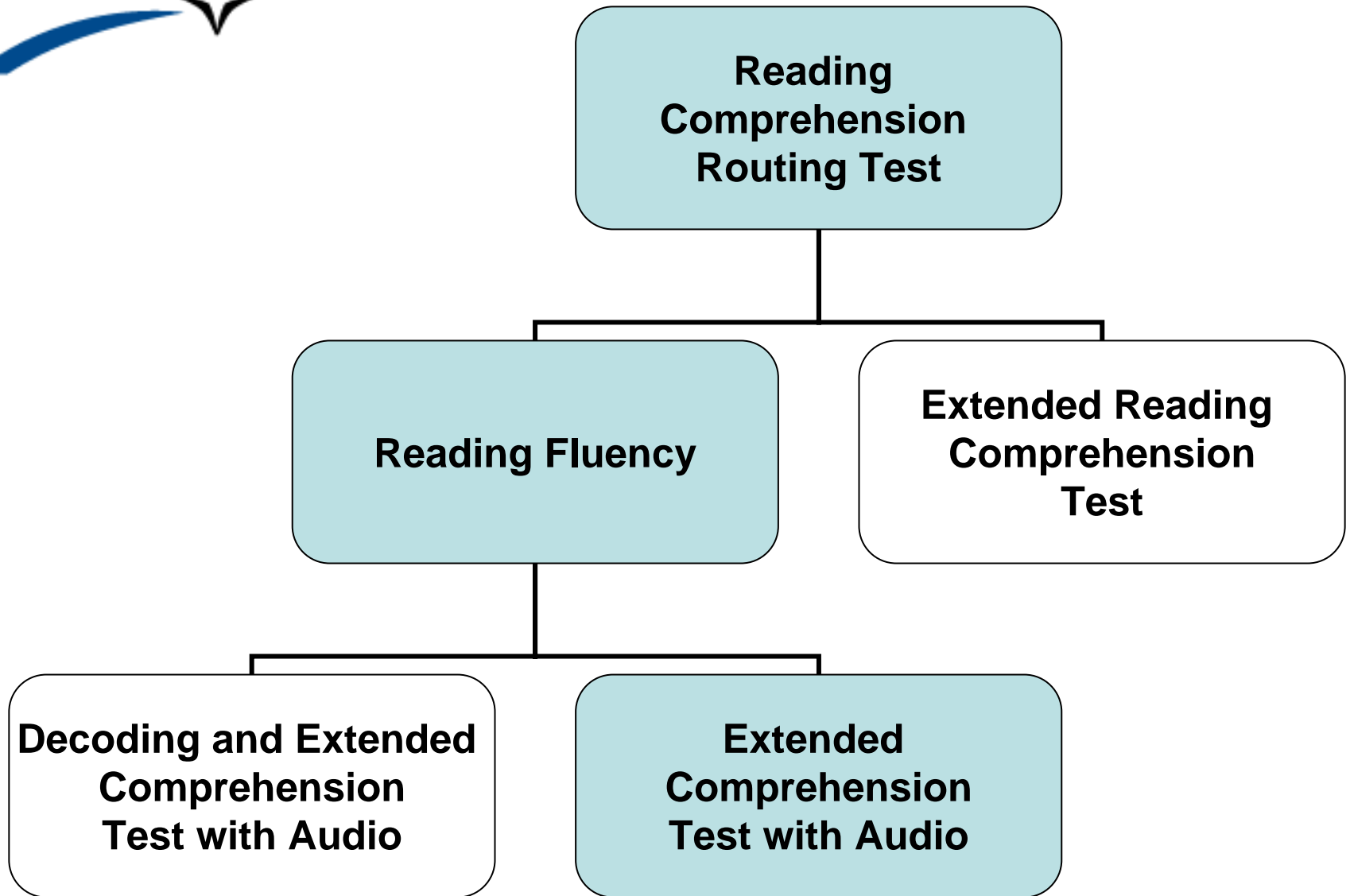
Multi-stage Testing: Process

- A first stage test will route a test taker with disabilities to one of several second stage tests. The second stage tests will be matched to the examinees' ability levels; estimated during the first stage of testing.
- One of the second stage tests could possibly provide an alternate assessment for the very low ability examinees (e.g., those who score at chance level on the grade level assessment).











Lots of Questions

- What proportion of students would be routed to anything other than the extended RC test?
 - More than 2 percent?
- Is routing test accurate?
 - Are there some students who do well on more difficult items but chance level on easy items?
 - Are there some students who respond at chance level on first 20 items and then start to show their true skills?



More Questions

- Should the Component test (decoding, fluency, and comprehension with audio) be considered a modified assessment?
- Can total scores on the Component test be reported on the same scale as then Integrated standard grade level assessment?
- Can scores on the audio extended and the standard extended be reported on the same scale?



More Questions

- How should we weight different measures?
- Can teachers predict which students would be routed and which ones would take the standard route?
- Can we create the routing and extended test using the same number of items currently on assessments of grade level standards?



Some Answers

Q: What portion of students would be routed to anything other than the extended RC test?

A: Approximately 97% of all students would take the standard assessment (integrated) if routing was based on chance level responses

- 98-99% of students without disabilities respond at chance level on current (integrated) state reading assessments
- 20% of students with learning disabilities respond at chance level on current (integrated) state reading assessments (about 2%)



More answers

Q: Can scores on the audio extended and the standard extended be reported on the same scale?

A: Yes

To answer this the DARA project conducted a series of research studies focused on the comparison of tests and test items that are read aloud and those that are not

- Differential Item Functioning
- Factor Analysis
- Score gains from read aloud for students with and without RLD



DARA Research Questions (Year 2)

- 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



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



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
- 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



Results of DIF and Factor Analysis

- DARA DIF and factor analysis studies indicate that test measures the same underlying construct for students with reading-based LD and students without disabilities when taken with and without a read aloud test change.
- Other research studies showed similar finding but one study found more DIF when reading tests were read aloud than when math tests were read aloud.
- When DIF exists the direction of the DIF is equally divided between items that are more difficult for the group that received read aloud and items that are easier for the group that received read aloud.



Research Answers (Year 2)

- Reading test items that are read aloud (via an individual CD player) are differentially easier for students with reading based LD compared to students without disabilities
- Reading tests and test items taken with and without read aloud:
 - Have the same factor structure
 - Have little evidence of differential item performance
 - But do not predict reading comprehension in the same way for students with and without disabilities
- However, a component assessment of reading (comprehension read aloud + reading fluency) predicts reading comprehension as well for students with RLD as the integrated comprehension measure does for students without disabilities



Next Steps

- Address the remaining questions
 - Is a single routing test accurate?
 - How many items should be on the routing test?
 - How should we weight different measures?
 - Can total scores for each assessment be reported on the same scale?
 - Can we create the routing and extended test using the same number of items currently on a standard state assessments?
- Studies will simulate the DARA proposed multi-stage assessment using both operational test data and data from experimental administration



Questions?

Cara Cahalan Laitusis, ccahalan@ets.org

or

visit www.ets.org/dara