Innovations in Measurement in Early Childhood Early Literacy

COLLEGE OF EDUCATION
+ HUMAN DEVELOPMENT

UNIVERSITY OF MINNESOTA

Today's Agenda

- EC RtI, CRtIEC, and the history of IGDIs
- Validity framework for developing second generation IGDIs
  - Construct Identification to Support Early Literacy Measurement
  - Task Development and Item Analyses in Innovative Measures of Early Literacy
  - Scaling Individual Growth and Development Indicators
  - Standard Setting with Innovative Measures of Early Literacy

EC RtI, CRtIEC, and the history of IGDIs

Early Childhood Response to Intervention (EC RtI)

- RTI: a systematic problem-solving process designed to
  - provide students with a level of instructional intensity matched to their demonstrated response to intervention.
  - provide a data-based method for evaluating the effectiveness of instructional approaches and changing/improving them.
- RTI is intended to reduce the need for special education by improving and providing services early.
- RTI services are individualized and based on evidence-based strategies.
- RtI services assume a high quality of "general" intervention, and add resources and services as needed.

Center for Response to Intervention in Early Childhood (CRtIEC)

- Mission: To conduct research and provide resources to support the application of RTI in Early Childhood Education.
- Objectives
  - Develop and validate
    - Tier 2 and 3 interventions for preschool language and early literacy skills
    - Identification and progress monitoring measures linked to these interventions
  - Evaluate Efficacy
    - Large scale studies to evaluate the efficacy of the Tier 2 and Tier 3 interventions
    - Supplementary studies of related issues
  - Provide national leadership on development and implementation of RTI for young children
  - Broadly disseminate findings to practitioners, researchers and policymakers

Why Early Literacy?

- Importance of early literacy to later academic success
- Identification of literacy problems is often too late
- Prekindergarten programs need evidence-based approaches for early identification and intervention in literacy and language
- These tools need to fit the reality of early childhood education settings

Individual Growth and Development Indicators (IGDIs): The History

- GOMs
  - 1) brief and easy to collect, 2) easily interpretable, 3) inexpensive, 4) related to important long-term outcomes and 5) repeatable.
- Technically adequate, high practical utility and social validity
- Design flaws that limit utility in an EC RtI framework

Construct Identification to Support Early Literacy Measurement

Tracy Bradfield

Wilson's Approach to Measure Development

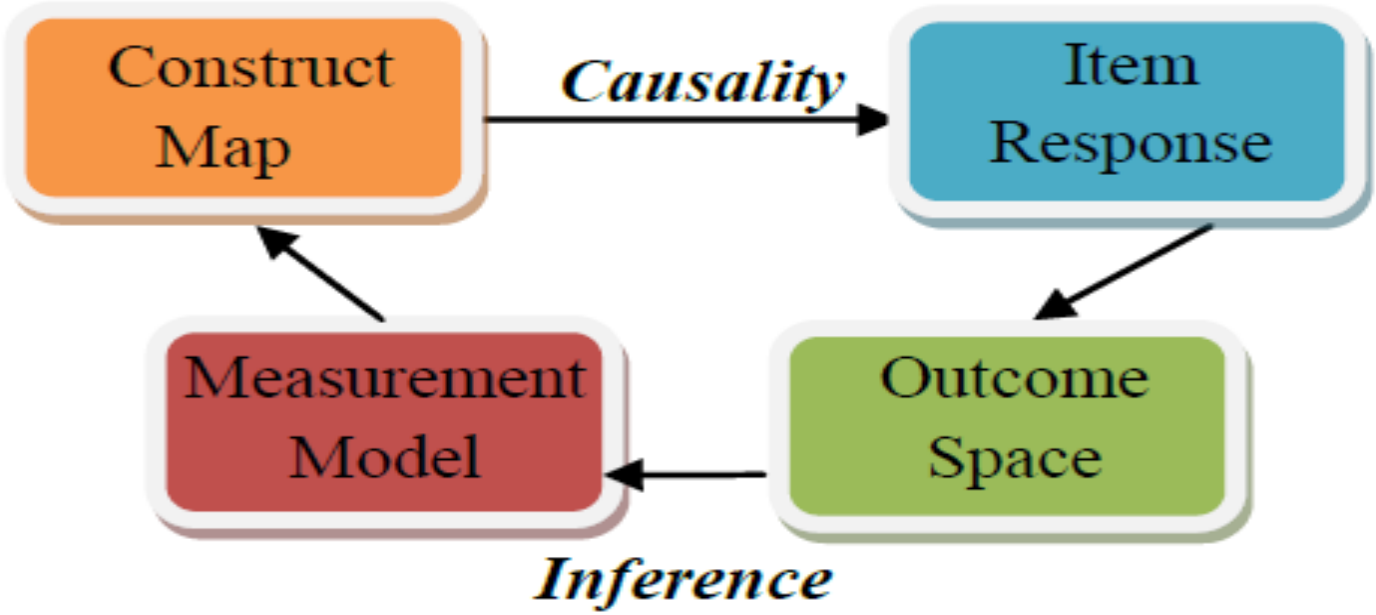


Figure 1. The four building blocks of an item response model approach to measurement construction (Wilson, 2005).

Building Blocks

- Construct Definition
  - A simple form: More or less, high to low
- Item Development
  - Realizations of the construct
- Outcome Space
  - Aspect of response we value – how to score
- Measurement Model
  - How we related scores to constructs

Construct Map

- Explains the construct; interpretation guide.
- Enables design of items that will lead children to give responses that inform important levels of the construct map; identify relevant item feature.
- Provides criterion to analyze responses regarding degree of consistency with construct map.
- Item selection or retention should be based on informed professional judgment.

Construct Map Development Process

- Extensive literature reviews process
  - Theoretical
    - Convergence of information
    - Review and approval by field experts
  - Measure Review
    - To inform identification of best forms of the construct
    - Included consideration of properties of GOMs and considerations of preschool assessment.

Oral Language Construct

- Operational Definition:
  - Ability to use words to communicate ideas and thoughts and to use language as a tool to communicate to others.
  - Expressive language: the use of words to express meaning.
  - Receptive language: ability to listen, process, and understand the meaning of spoken words.
- Form: Vocabulary knowledge including formats that tap both a child's receptive and expressive vocabulary.

Phonological Awareness Construct

- Operational Definition: Phonological awareness is the ability to detect and manipulate the sound structure of words independent of their meanings.

- Form(s): rhyming, alliteration, syllable segmenting, and sound blending

Alphabetic Knowledge Construct

- Operational Definition*:* knowledge about the names and sounds of letters.

- Form(s): a child's ability to demonstrate knowledge of the names and corresponding sounds of the 26 letters of the alphabet

Comprehension Construct

- Operational Definitions:
  - Text comprehension: the recognition of pictures and symbols and the ability to interpret and infer meaning from what is seen.
  - Listening comprehension: the ability to understand and interpret what is spoken aloud and infer meaning from what is heard.
- Form(s): Asking children to choose the picture associated with verbally presented stimuli, story retell with and without verbal cloze prompts, and asking a child questions about a story after it had been read to him.

Measure Development

Alisha Wackerle-Hollman

Research Process

- Extensive Literature Reviews
- Robust analyses of component skills for each early literacy area: Alphabetic Principle, Oral Language, Phonemic Awareness and Comprehension
- Research Design and Pilot implementation
- Phase 1 trials
- Finding the best candidates for further testing– evaluating the measures
- Phase 2
- Refinement and Revisions

Oral Language

- "Ability to use words to communicate ideas and thoughts and to use language as a tool to communicate to others."
- Link between vocabulary development and later reading
- There are two major categories:
  - **Expressive Language** (Ability to use words to convey meaning)
  - **Receptive Language** (Ability to listen, process, and understand the meaning of spoken words)

Oral Language Measures

- Picture Naming
- Definitional Vocabulary with Pictures
- Definitional Vocabulary without Pictures
- Motor Instruction
- Point to Picture
- Which One Doesn't Belong

Expressive Language Measures

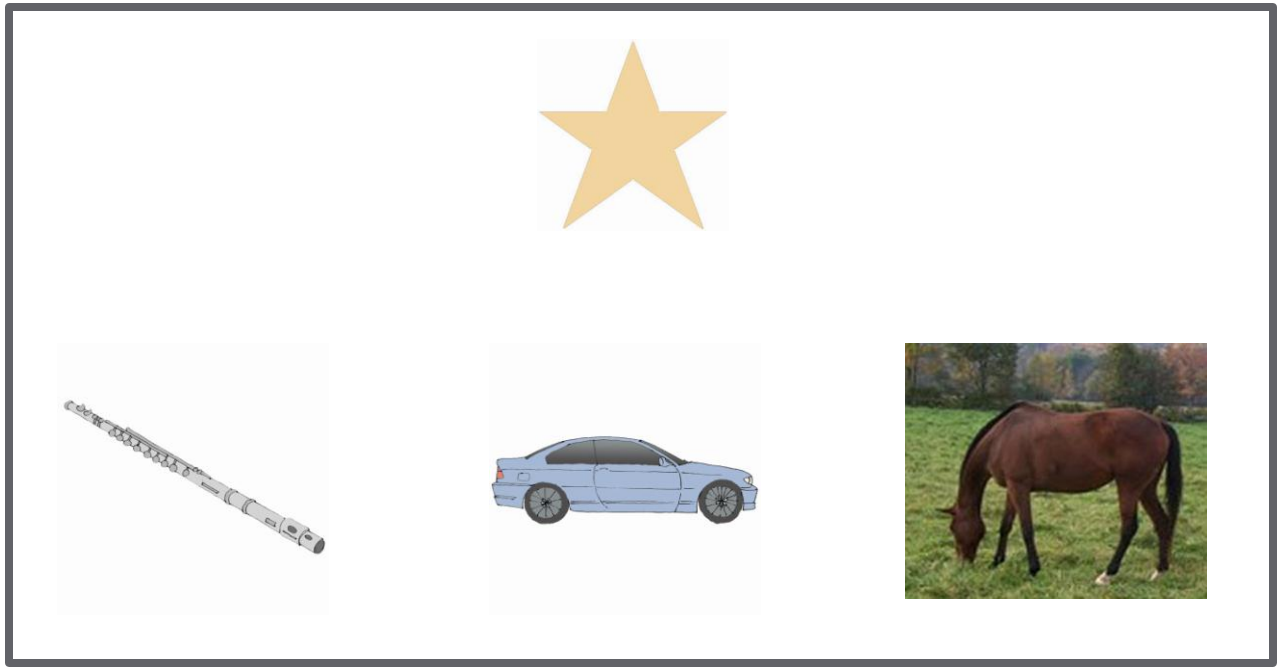
- Picture Naming

Phonological Awareness

- Phonological awareness is the ability to detect and manipulate the sound structure of words independent of their meanings. (Phillips, Clancy-Menchetti, Lonigan, 2008)
- Phonological awareness is one of the best determinates of successful early reading acquisition. (Foy & Mann, 2006)
- Four measures were created, each lasting 1 to 2 minutes:
  - Rhyming,
  - Alliteration,
  - Sound Blending
  - Syllable Segmenting.

Phonological Awareness Measures

- Identify picture that rhymes with target picture; 2 minutes



PA Measure in action

Alphabetic Knowledge

- *Alphabetic knowledge* is knowledge about the names and sounds of letters (McBride-Chang, 1999).
- Five potential IGDIs were developed that we believe spanned the ability range within the domain of alphabetic principle.
  - Letter Orientation
  - Letter Naming
  - Letter Identification
  - Sound Identification
  - Sound Naming

Alphabet Knowledge Measures

Point to the letter.

Comprehension

- *Comprehension* in preschoolers consists of two parts:
  - *Text comprehension* is the ability to understand and interpret text, including pictures and symbols (Storch & Whitehurst, 2002; Dunst, Trivette, Masiello, Roper, & Robyak, 2006).
  - *Listening comprehension* is the ability to understand and interpret spoken language at multiple levels (Dickinson & Smith, 1994; Skarakis-Doyle, Dempsey, & Lee, 2008).
- An important component of comprehension is *inferencing*, or filling in information that was not actually seen or heard (Kendeou, Bohn-Gettler, White, & Van den Broek, 2008).

Comprehension Measures

- Sequencing
- Sentence Comprehension
- Picture Comprehension
- Story Comprehension
- Story Titles Task
- Which One Doesn't Belong

Comprehension Measures

Item level Revisions

- Cleaning Items
- Item level functions
  - Rasch Output Values
    - How is each item contributing to the test?
  - Construct Irrelevant Features (CIF)
    - What characteristics of each item provide information to the student?
    - What information distracts the student from the intended content?
    - What features of the items are malleable?

Construct Irrelevant Features

Rasch output indicated this item was problematic because it was very difficult at a value of 2.41

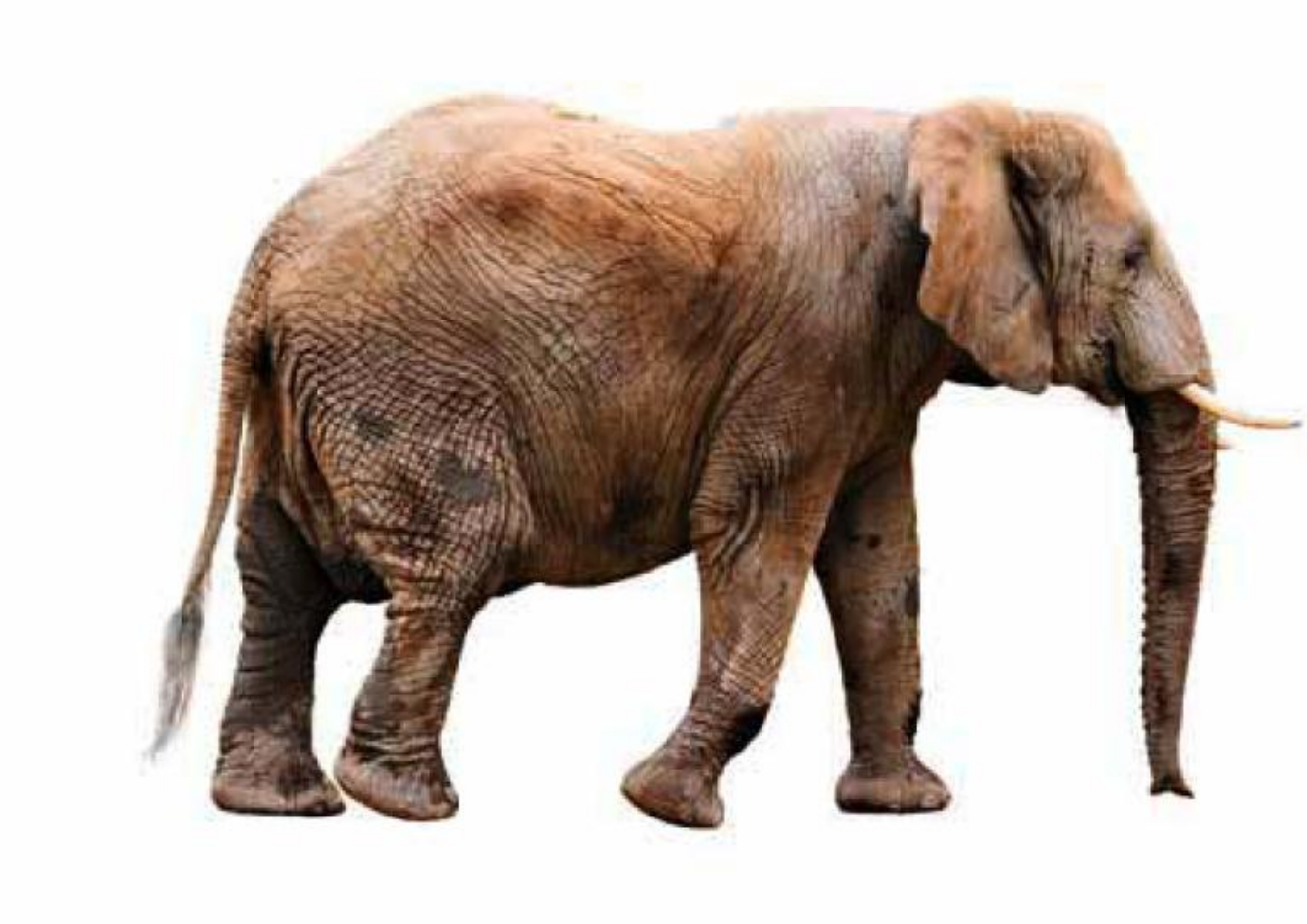
Example: Poorly Functioning Item – Def Vocab

Construct Irrelevant Features

Not a real elephant

Elephants are big but this one is actually small.

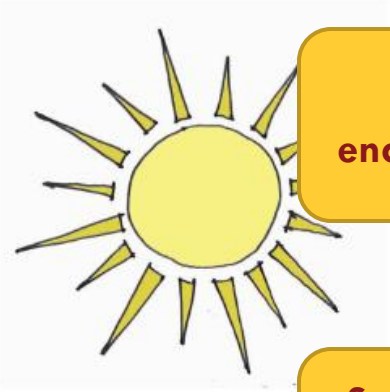
Example: Revision of Item



Example: Poorly Functioning Rhyming Item

Construct Irrelevant Features

Word content, color content and image clarity all might contribute to a response.



Some are enclosed, others are not

Some are real images others are not

Example: Revised Item

"Bat, Doll"

"Bat, Cat"

Summary

- Phase 2 testing yielding important information about item revisions, to be tested as more items are created and included in the catalog of items available for each task.
- Phase 2 results will be scaled using Rasch modeling to determine if student performance is adequately matched with items at appropriate difficulty levels.
- We will continue to refine and reduce the number of measures in the IGDI 2.0 cadre.

Scaling

Anthony D. Albano

Purpose

- Identify specific measures, and items within the measures, which function well as indicators of growth and development with preschool children
- Examine measures in terms of:
  - scale reliability
  - correlations with criterion measures
  - standard errors of parameter estimates
  - overlap between item and person distributions

Challenges

- Items for each measure could not be administered in one sitting
- Students saw different items at each of three waves (Fall, Winter, Spring)
- Data collection design: items and waves partially crossed with students, nested within classrooms
- A given item was only seen by 100 to 200 students at a given wave

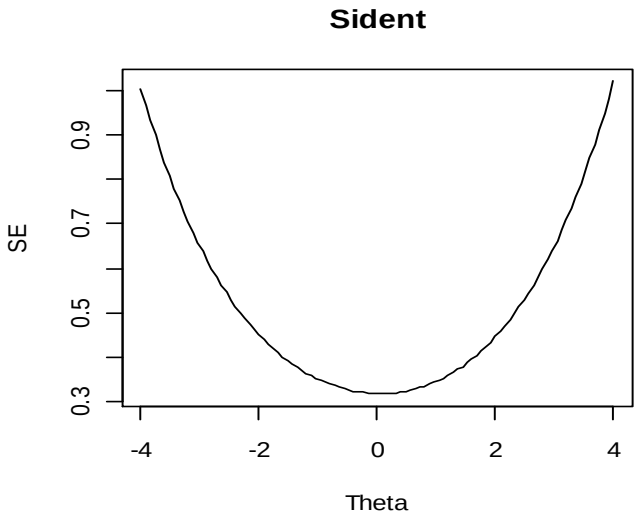
Calibration

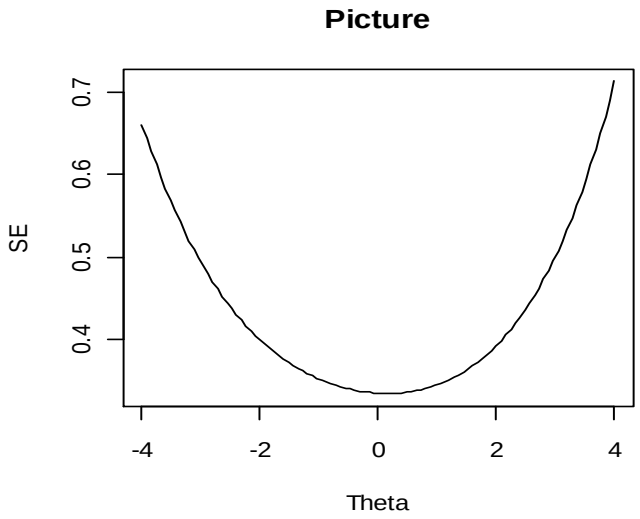
- A Rasch model was fit, via Winsteps, with the theta scale centered around the average item difficulty
- The initial scaling involved calibration of items and people at wave 1
- Ability estimates were obtained at waves 2 and 3 by fixing item parameters at wave 1 values

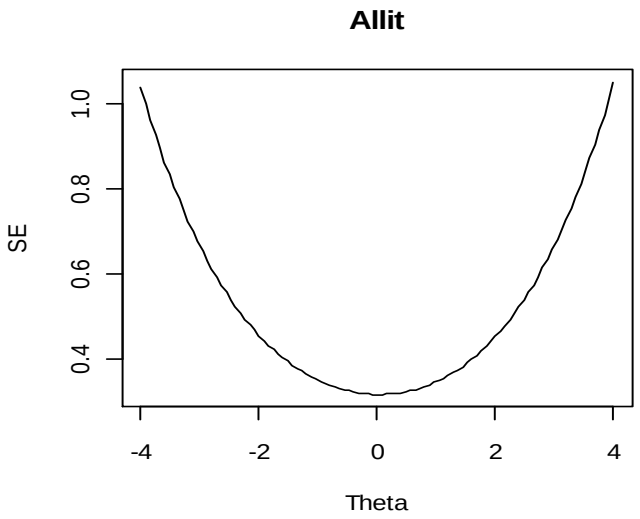
Reliabilities and Criterion Correlations

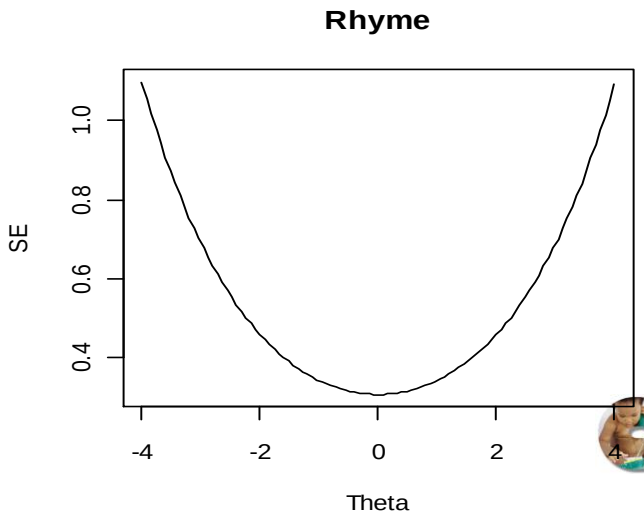
Measure	Domain	r	Criterion	N
Sound ID	Alpha Knowledge	.60	.56	662
Letter	Alpha Knowledge	.70	.53	663
Alliteration	Phono Awareness	.42	.41	681
Rhyming	Phono Awareness	.56	.43	706
Picture	Oral language	.80	.68	695
Which One	Oral language	.72	.49	637
Vocab	Oral language	.57	.61	654

Standard Error Functions

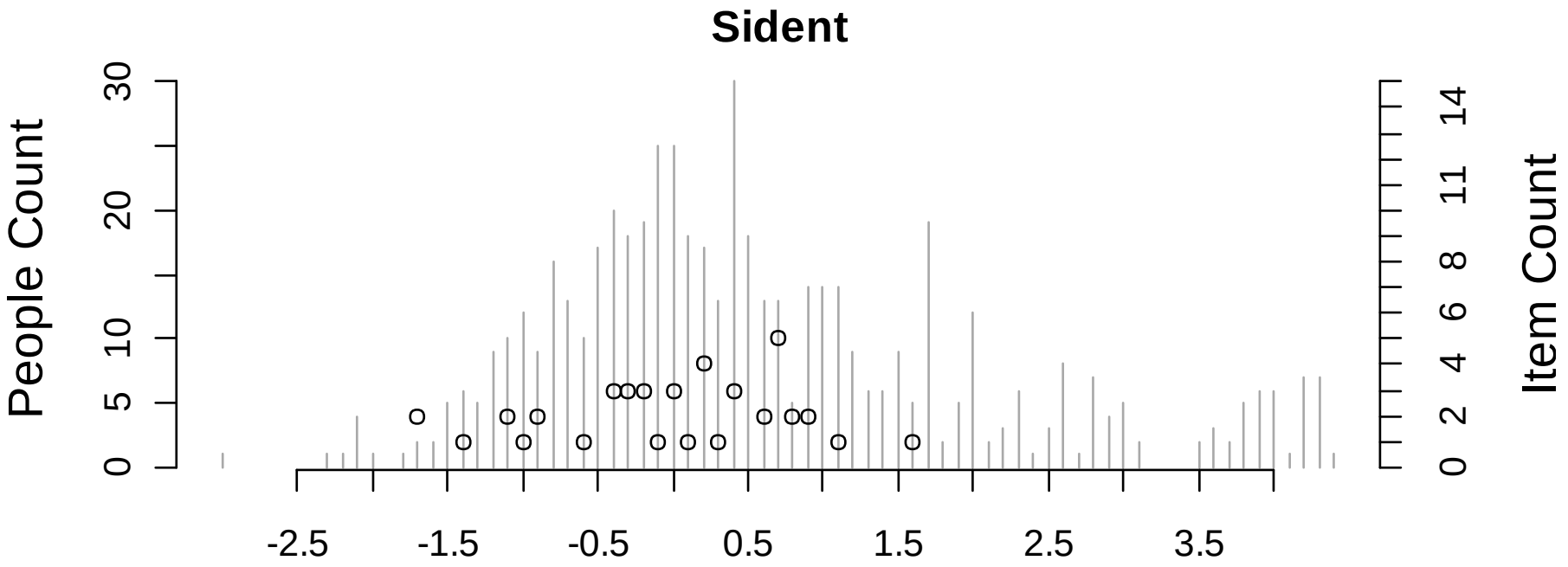




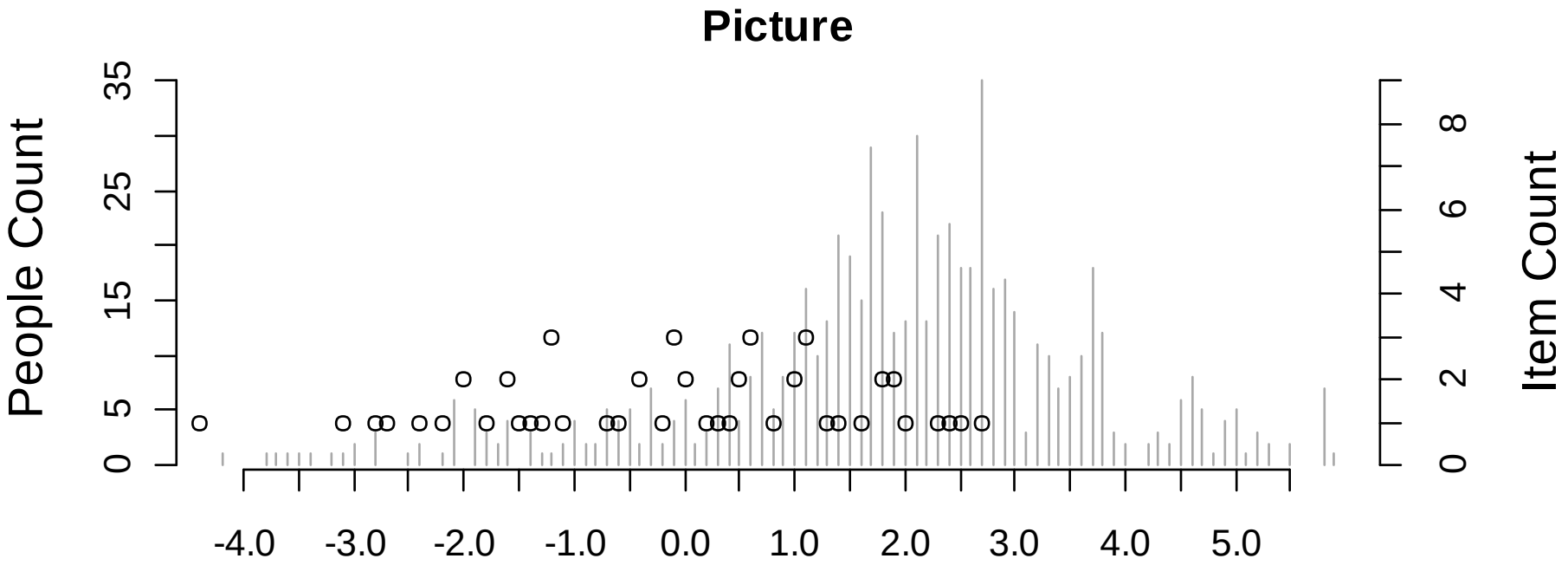




Item-Person Distributions



Item-Person Distributions



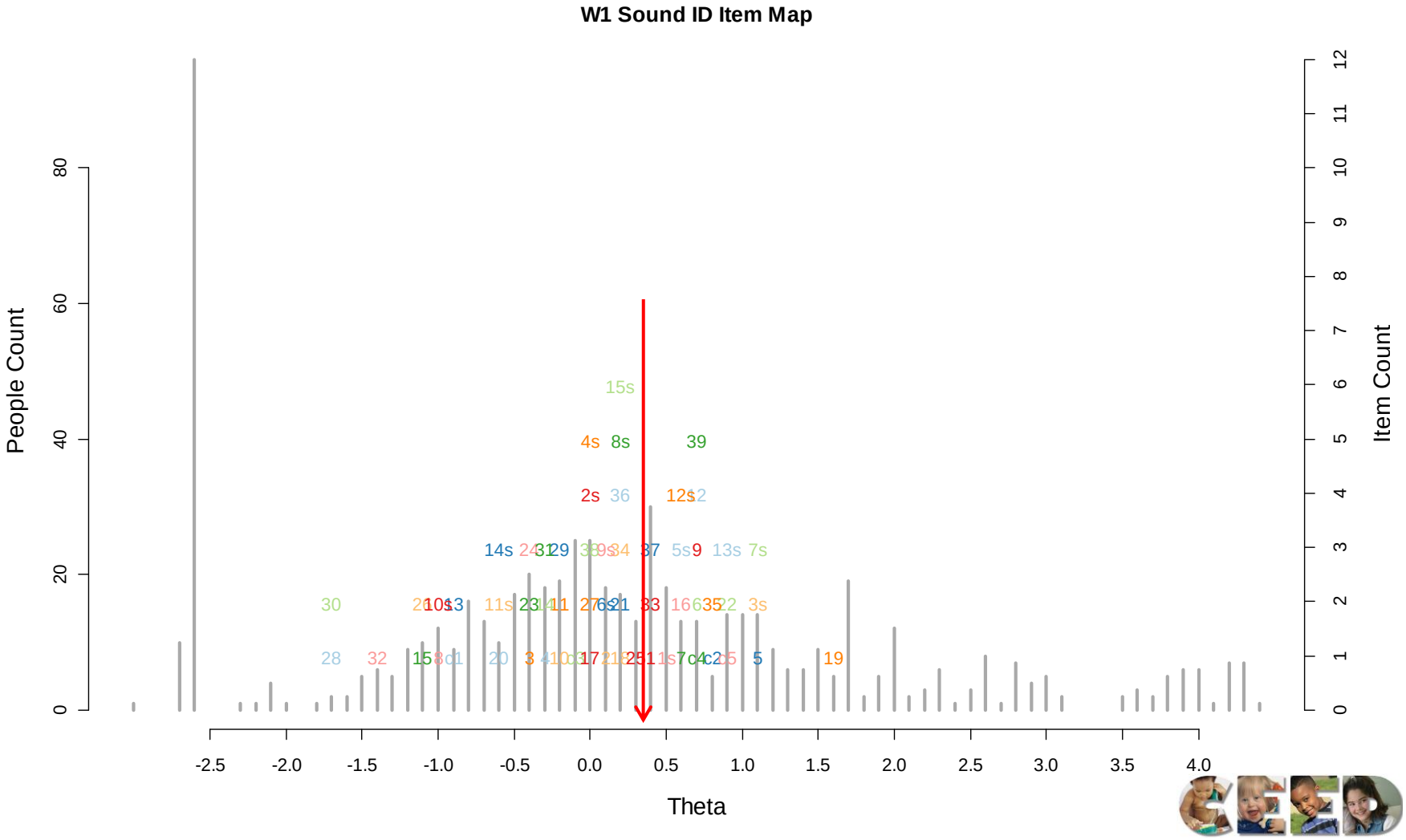
Decisions

- Based on these results, Sound ID and Picture Naming were selected for further development
- The final criterion - feasibility of future item development
- Anticipated longevity of the measure

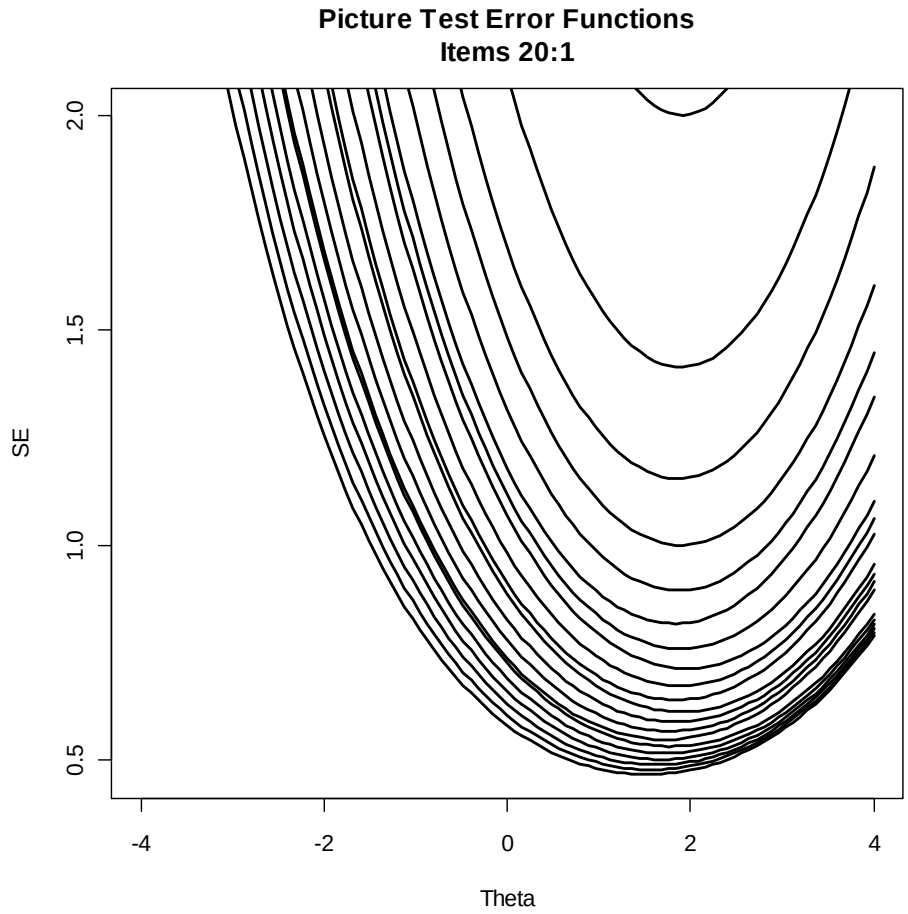
Identification Bundles

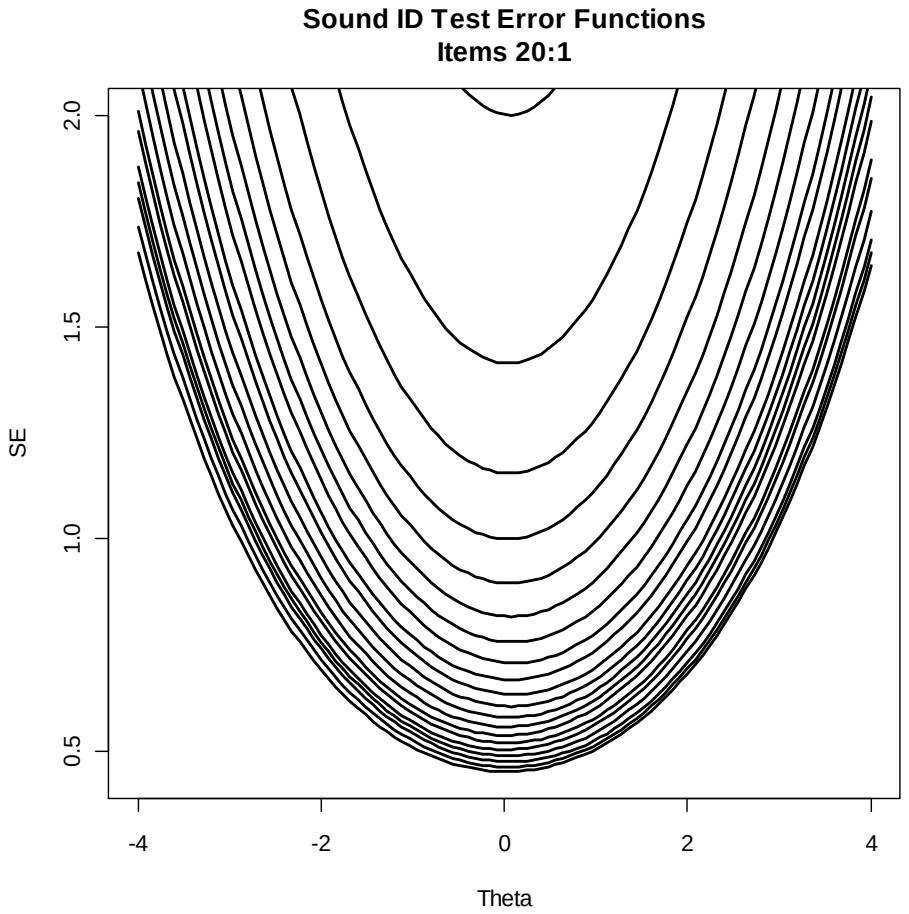
- Reduction of final measures to ID forms
- Identify cut points
- Amass high quality items around the cut

More Item Mapping



More Standard Error Functions





Standard Setting

Michael C. Rodriguez

Standard Setting

- Function of RTI – identification
- Set cut scores to support placement in Tiers 2 and 3
- Contrasting Groups, based on teacher nomination (teacher judgment)

Kansas Experience (Poggio, 1984)

- the method is rather easily implemented;
- teachers report little difficulty in following what is to be done in Contrasting Groups;
- the public is both confused and tends to doubt the legitimacy of the standard when they (often) cannot understand the "statistical magic" which delivers the standard (commonly associated with more complex standard setting methods); and
- the method gives support to the idea that "teachers can already tell us who is competent."

Contrasting Groups Design

- Teachers complete a child performance survey
- To place children into a Tier level (1, 2, or 3), based on their understanding of the performance level from the tier level descriptors (TLDs)
- Children assessed on the IGDIs
- Used ROC analysis to select cut scores

Teacher Classification of Students

Oral Language Tier Placement		Alphabet Knowledge Placement 1	2	3	OL Total
1	Count	269	102	22	393
	% of Total	46%	17%	4%	67%
2	Count	43	65	35	143
	% of Total	7%	11%	6%	**24**%
3	Count	5	18	28	51
	% of Total	1%	3%	5%	**9**%
AK Total	Count	317	185	85	587
	% of Total	54%	**32**%	**15**%	100%

ROC Analysis of Picture Naming

- The test statistic associated with a ROC analysis, the probability that a score for a randomly selected positive case (Tier 2/3) is lower than the score for a randomly selected negative case (Tier 1).
- The ROC analysis suggests successful classification based on PN, Area = .782, $p < .001$.

Sensitivity

- the probability that the "test" says a child has limited language development or skills when in fact they do have the limitation (true positive).
- These are the children we hope to identify for potential placement in Tiers 2 or 3, as they are likely to benefit from such placement.

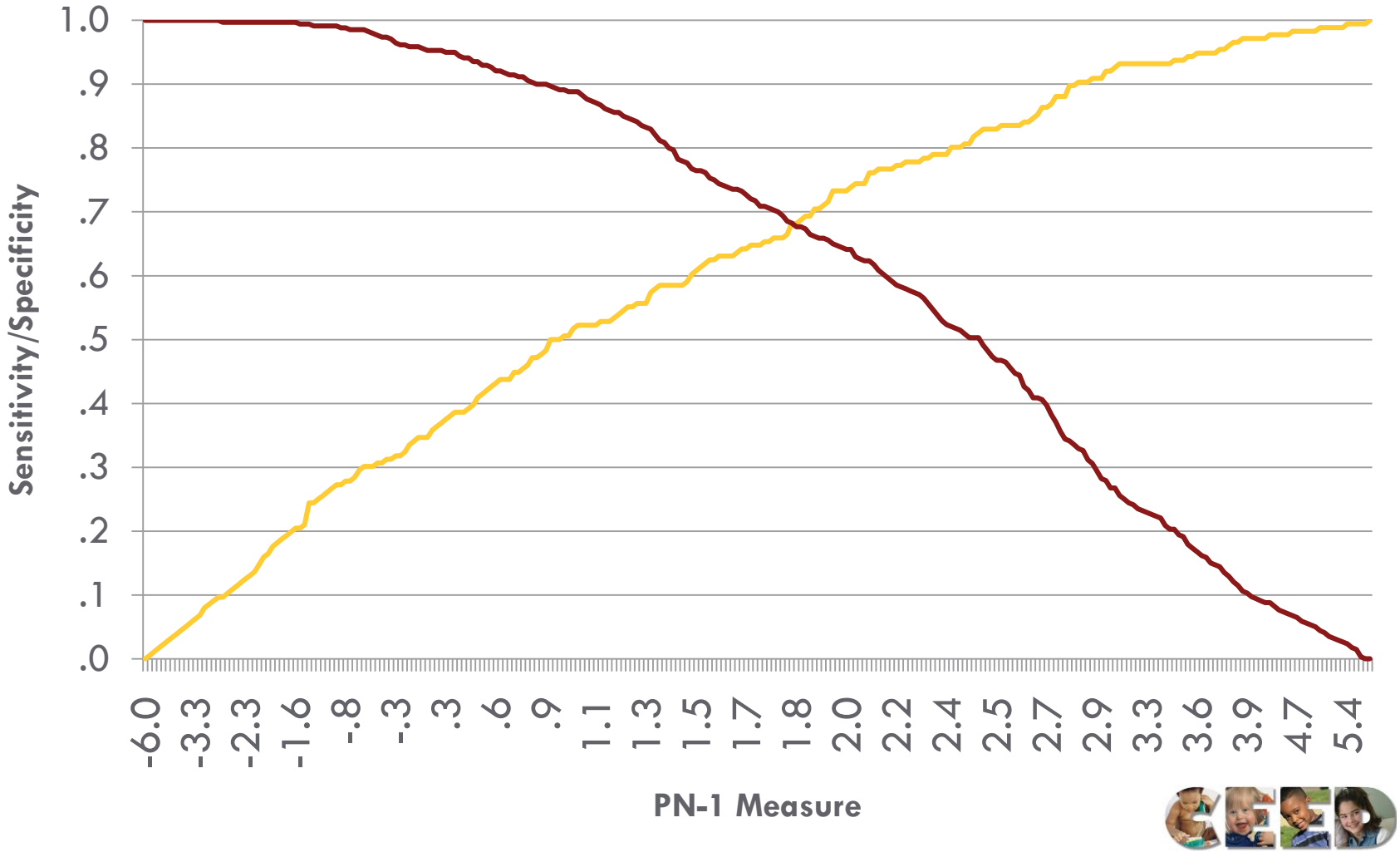
Specificity

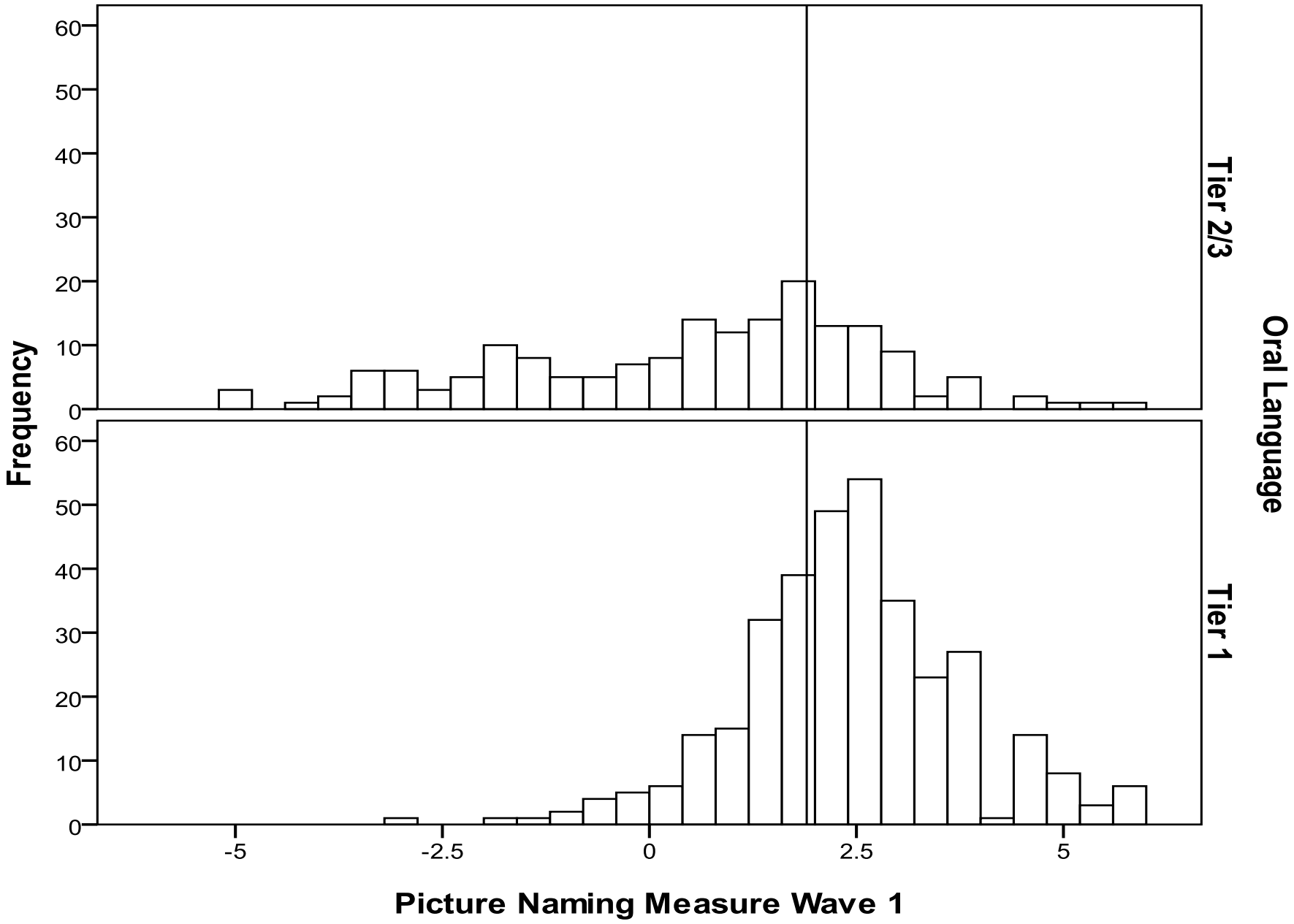
- the probability that the "test" says a child does not have limited language development when in fact they do not (true negative).
- These are the children we hope to NOT identify for potential placement, such that they remain in Tier 1, as they are not in need of Tier 2 or 3 intervention.

ROC Results for Picture Naming

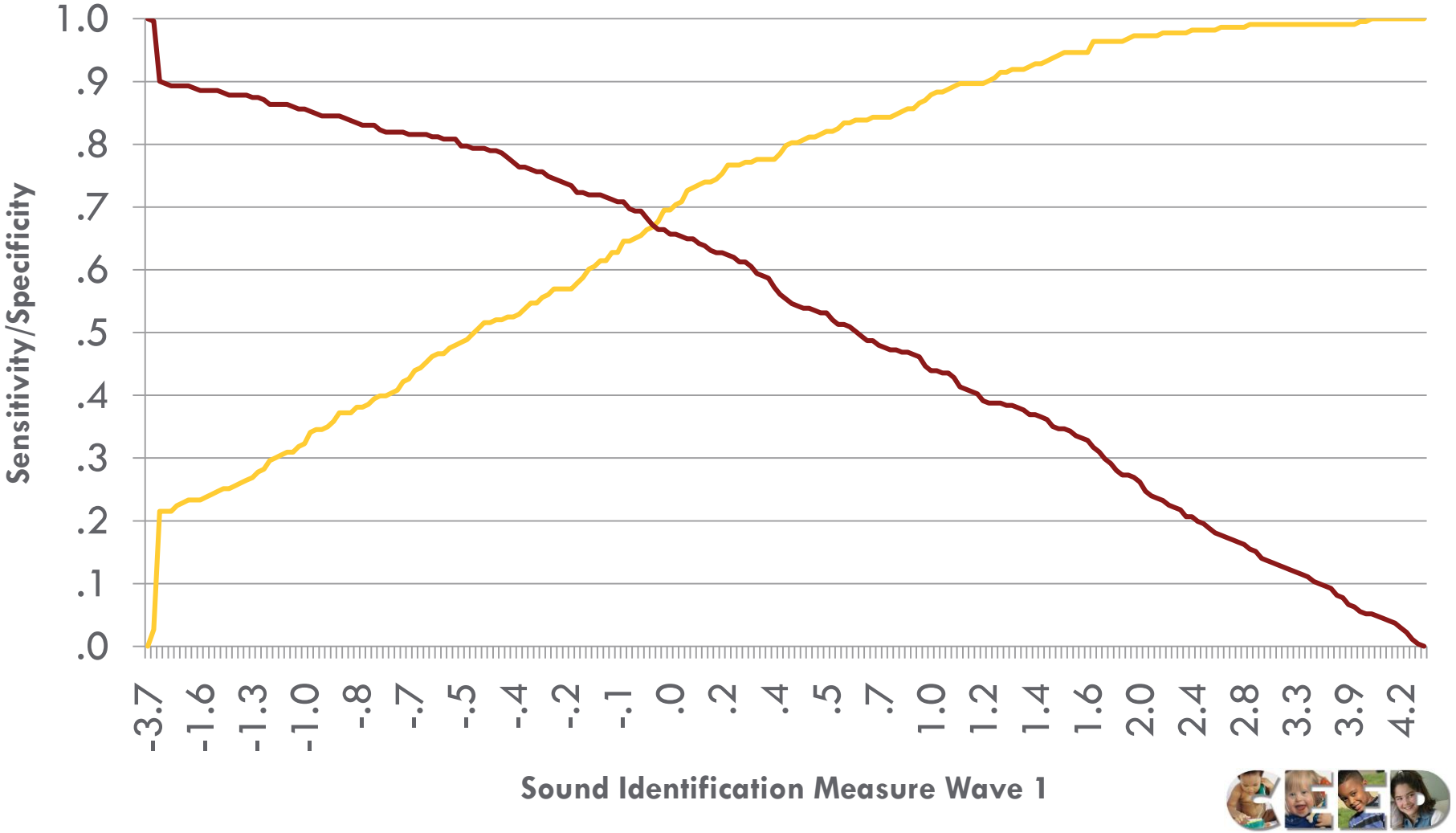
- The cut score that yields .70 sensitivity and .66 specificity is 1.90 logits.
- Although based on the ROC analysis results, the maximal value on both metrics is about .68, which is associated with a score of 1.81.

ROC Analysis for Picture Naming

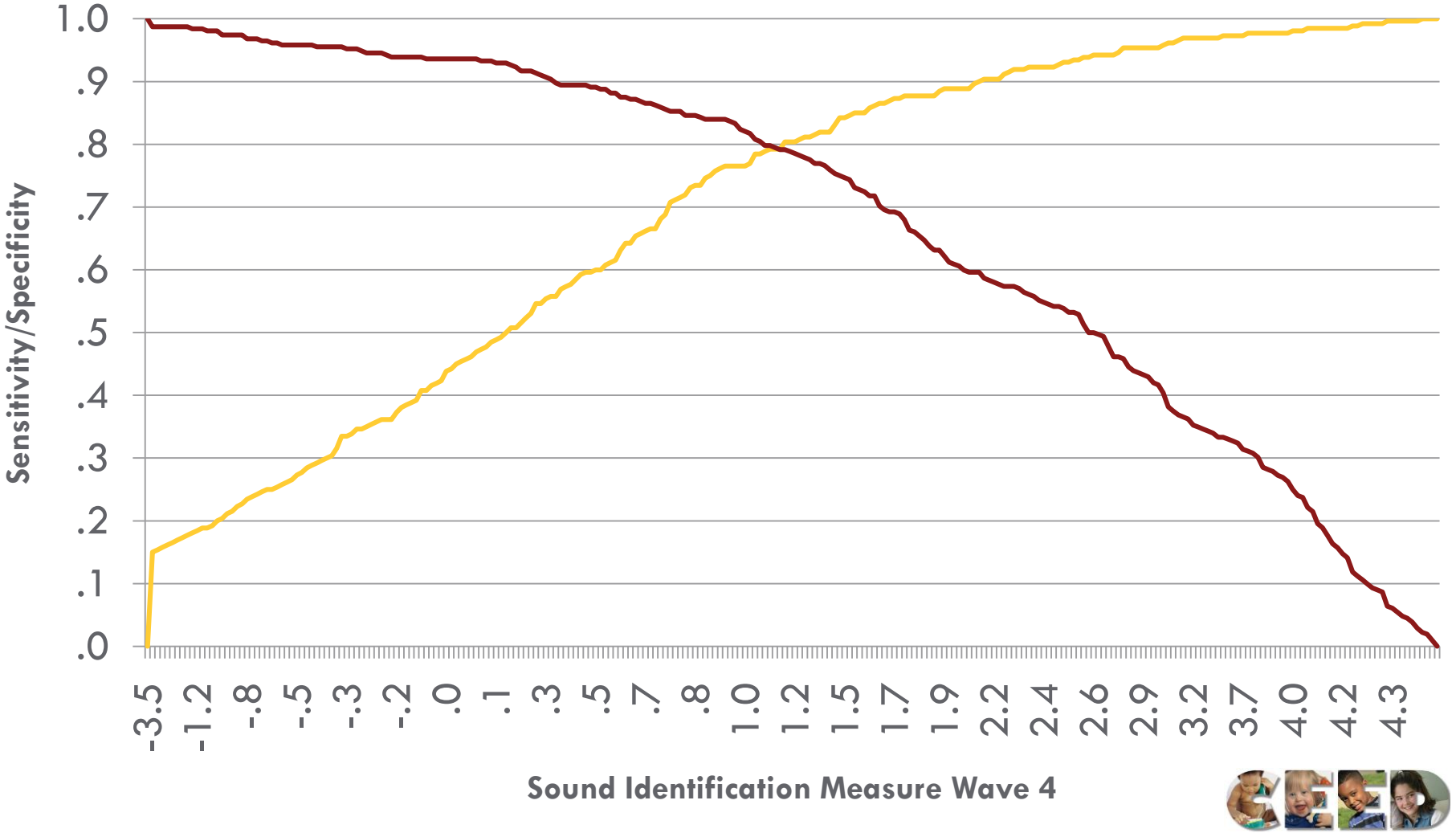



Frequency
60
50
40
30
20
10
0
Tier 2/3
Oral Language
60
50
40
30
20
10
0
Tier 1
-5
-2.5
0
2.5
5
Picture Naming Measure Wave 1

ROC Analysis for Alphabet Knowledge



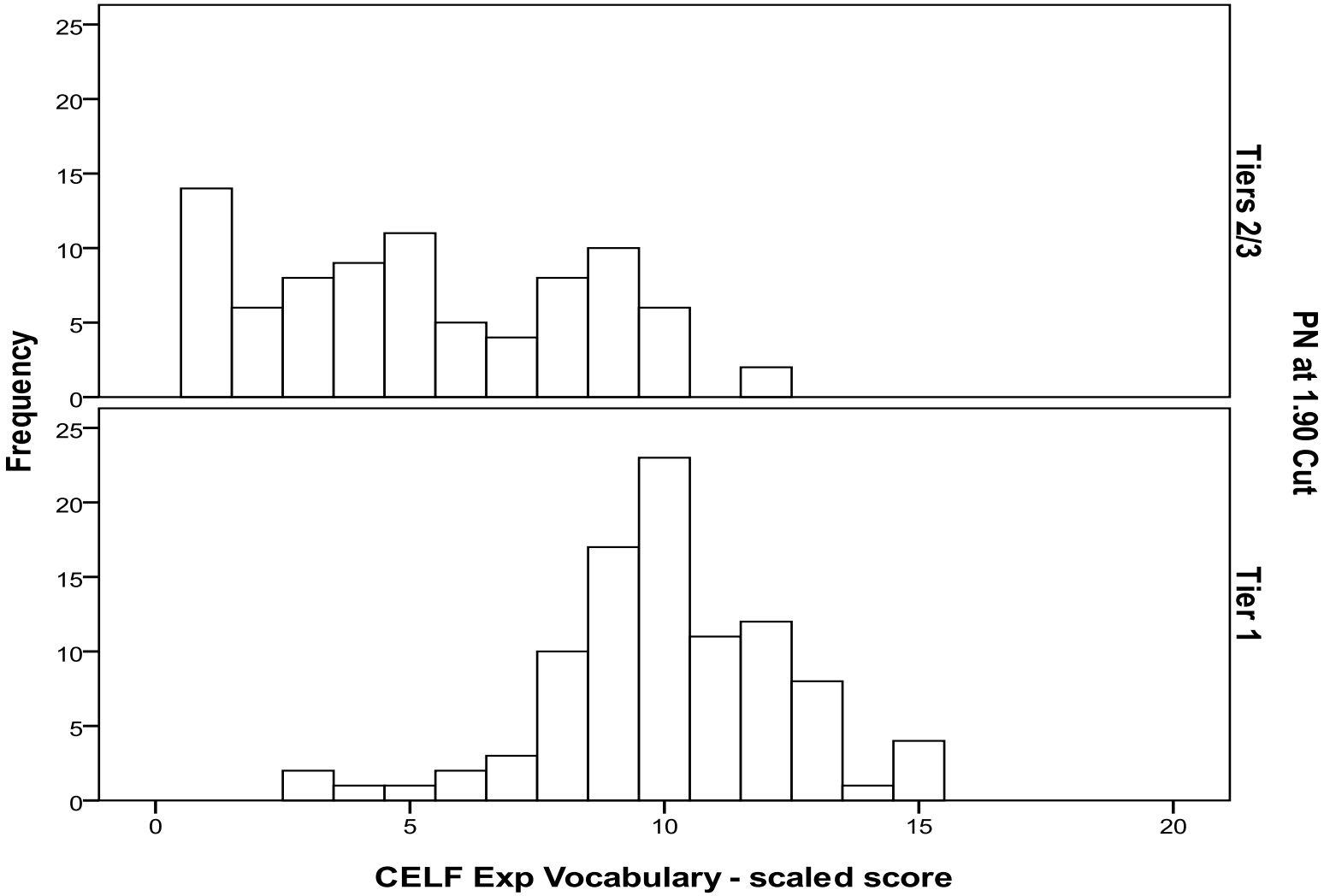
ROC Analysis for Alphabet Knowledge



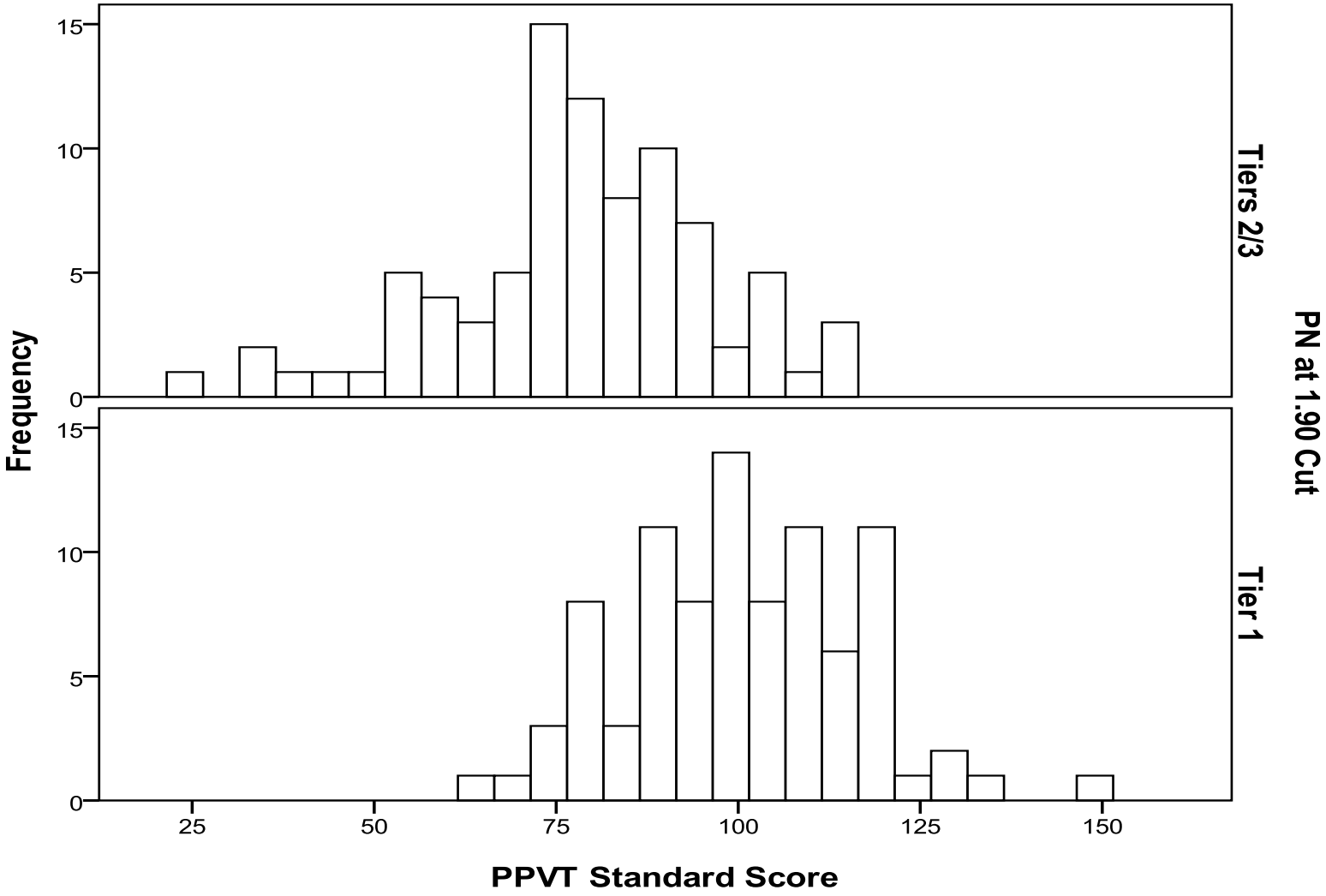
Validity Considerations

- An initial possibility is to assess Tier placement with a number of IGDIs.
- The amount of time a child spends within a Tier (as an indication of responsiveness to intervention)
- Reassess following transitions out of a Tier, providing follow-up evidence about performance following intervention.

Performance on Other Measures



Performance on Other Measures



Thank You

www.crtiec.org