Monitoring Progress On Progress Monitoring: Recent Innovations In The Design And Evaluation Of Individual Growth And Development Indicators

Presented at the Conference on Research Innovations in Early Intervention, February 26, 2010

Today's Presenters

- Scott McConnell, Center for Early Education and Development, University of Minnesota
- Kristen Missall, Department of Educational Psychology, University of Iowa
- Michael Rodriguez, Department of Educational Psychology, University of Minnesota
- Alisha Wackerle-Hollman, Center for Early Education and Development, University of Minnesota

IGDIs – A Brief Introduction

- Individual Growth and Development Indicators
  - **General Outcome Measures** of development for young children
    - Based on work of Deno and others
  - Intended to be time- and cost-efficient, repeatable measures of child performance in one or more developmental domains
  - Designed to provide estimates of current status and growth over time
- Psychometrically rigorous
  - Reliability
  - Validity
  - Sensitivity
  - Screening, diagnosis, and growth
- Across age groups and investigators
  - Infant and toddler measures
  - Preschool measures
  - Early Elementary measures

IGDIs – A Brief Timeline

Pre-Formal Period

1977 Deno & Mirkin, Data-based program modification

1982 Deno, Mirkin & Chaing

1994 Priest, Spicuzza, McEvoy and McConnell early effort to develop growth measures for social interaction

IGDIs – A Brief Timeline

Initial Efforts

1996-2001 Early Childhood Research Institute on Measuring Growth and Development

- Infant IGDIs
- Preschool IGDIs
  - Picture Naming
  - Rhyming
  - Alliteration
  - Blending
  - [others]
- Dibels

~2002 Online data systems – *Get it, Got it, Go!*, Juniper Gardens IGDIs, Dibels Data System

IGDIs – A Brief Timeline

~2003 "Viral" growth in application

- Infant measures – ECI and Early Head Start
- Preschool measures and Early Reading First
- Dibels and Reading First/NCLB

>2003 Other investigators, new measures and applications in "old" and "new" places

2008 Early Childhood Research Institute on Response to Intervention in Early Childhood

- Formal adaptation to requirements of Response to Intervention

New cadre of IGDIs

- Old IGDIs
  - Picture Naming
  - Rhyming
  - Alliteration
- New IGDIs
  - Picture Naming
  - Definitional Vocabulary
  - Which One Doesn't Belong
  - Letter Orientation
  - Sound Identification
  - Rhyming
  - Alliteration
  - Sound Blending

IGDIs Now

- Can we leverage what we've done, learned, and gathered?
- Can we improve on existing approaches?
  - Problems of sensitivity, esp for younger and lower-performing children
  - Restricted reliability
  - Need for "benchmarks," standards, and applications

If your job is bolting the engine as the rocket launches, you get an outfit…

Today's Discussion

- *Can data provided by practitioners, from the assessment of children in early childhood programs throughout the country, be used to set normative standards for interpretation?*
- *In what ways do emerging psychometric analyses aid in measure development and evaluation?*
- *How should we develop new assessment formats?*
- *Are we there yet?*

CAN FIELD DATA BE USED TO SET NORMATIVE STANDARDS FOR INTERPRETATION?

Using *Get It, Got It, Go!* for Preschool EL-IGDI norms

Get it, Got it, Go!

- *Get it, Got it, Go!* is a website for Preschool EL-IGDIs (Picture Naming, Rhyming, Alliteration; http://ggg.umn.edu or http://www.getgotgo.net)
- *Get it* – get materials and information about administration
- *Got it* – enter data and information about individual children, get score reports, generate graphical reports about individual and group performance
- *Go!* – score and report interpretation, intervention resources

Got it!

- User registers with basic contact information
- User reads and accepts agreement about purpose of G3, copyright, privacy of data, secondary use of data, liability, and support
- User creates a profile and identifies…
  - District, school, user category (teacher, psychologists, paraprofessional), licenses held, type of IGDI training, years experience in education, highest degree
- User enters student demographic data
  - Basic demographic information + special education status (none to IFSP/IEP), special education category, race, home language, enrolled in programming as a child of a low-income family
- User enters individual IGDI data
  - Individual data can be grouped by class or other variable; both levels (individual and group) can be monitored, graphed, and analyzed

Mining Field Data for Analysis

- Currently, more than 125,000 cases entered into G3
- About 2 years ago we extracted G3 data for over 70,000 cases
  - Data in large-scale storage program; our request yielded .tsv files which were saved to Excel
  - We set criteria for cleaning data and deleted…
    - "junk files" – incomplete, made-up names
    - Limited age groups outside of 36 to 60 months old
    - Outlier data based on controlled studies
  - We selected variables to define our data set
    - "typical" = no IEP, no income eligibility, no ELL
  - Result was 32,615 data points on 7,358 "typical" preschoolers
    - 50.2% girls
    - PN = 12,466 data points, or 38.2% of total
    - RH = 10,336 data points, or 31.6%
    - AL = 9,813 data points, or 30.1%
  - Data exported from Excel to a stats program for analysis

Sample

- Children from 709 separate preschools (M = 11.2/school) and 463 school districts (M = 17.1/district) in 43 states (M = 184.8)
- Mean age 52.2 months for boys and 52.1 months for girls (SD = 6.13, range 36-60 mo)
- Ethnically diverse
  - 69.6% Caucasian, 16.1% African-American, 6.7% Hispanic, 6.3% Native American, 1.4% Asian-American
- 100% spoke English, 0% received special education, and 0% qualified as low income

Representation

- Sample examined for each EL-IGDI by racial/ethnic group at age 3 (36-48 mo) and age 4 (48-60 mo) for boys and girls
- Sample compared to US Census Bureau population parameters and percent differences examined
  - Some populations slightly over- (% > 0) or under-estimated (% < 0)

			Ethnicity[a]					
EL-IGDI	Age	Sex	Black	Asian/PacIs	White	Hispanic/Latino	American Indian	*n*
Picture Naming	3	F	227, 2.1%	18, -1.9%	858, -2.8%	51, -7.8%	63, 2.1%	1217
		M	225, 1.8%	14, -2.1%	873, -4.0%	42, -8.6%	38, 1.0%	1192
	4	F	665, -0.8%	82, -2.1%	3552, -1.8%	363, -6.1%	374, 3.2%	5036
		M	648, -1.2%	59, -2.3%	3510, -2.1%	429, -5.9%	375, 3.2%	5021
Rhyming	3	F	191, 3.4%	9, -2.2%	625, -3.3%	27, -8.4%	45, 2.0%	897
		M	191, 3.2%	8, -2.2%	638, -4.4%	20, -9.2%	26, 0.9%	883
	4	F	617, 0.7%	70, -1.6%	2936, 5.5%	338, -6.0%	332, 3.1%	4293
		M	612, 0.4%	55, -1.8%	2873, 2.5%	381, -6.0%	342, 3.4%	4263
Alliteration	3	F	180, 3.8%	10, -2.0%	565, -3.4%	27, -8.2%	45, 2.2%	827
		M	169, 2.9%	7, -2.3%	568, -5.1%	15, -9.4%	28, 1.2%	787
	4	F	580, -0.3%	70, -1.7%	2831, -3.8%	330, -5.7%	330, 3.5%	4141

Analyses

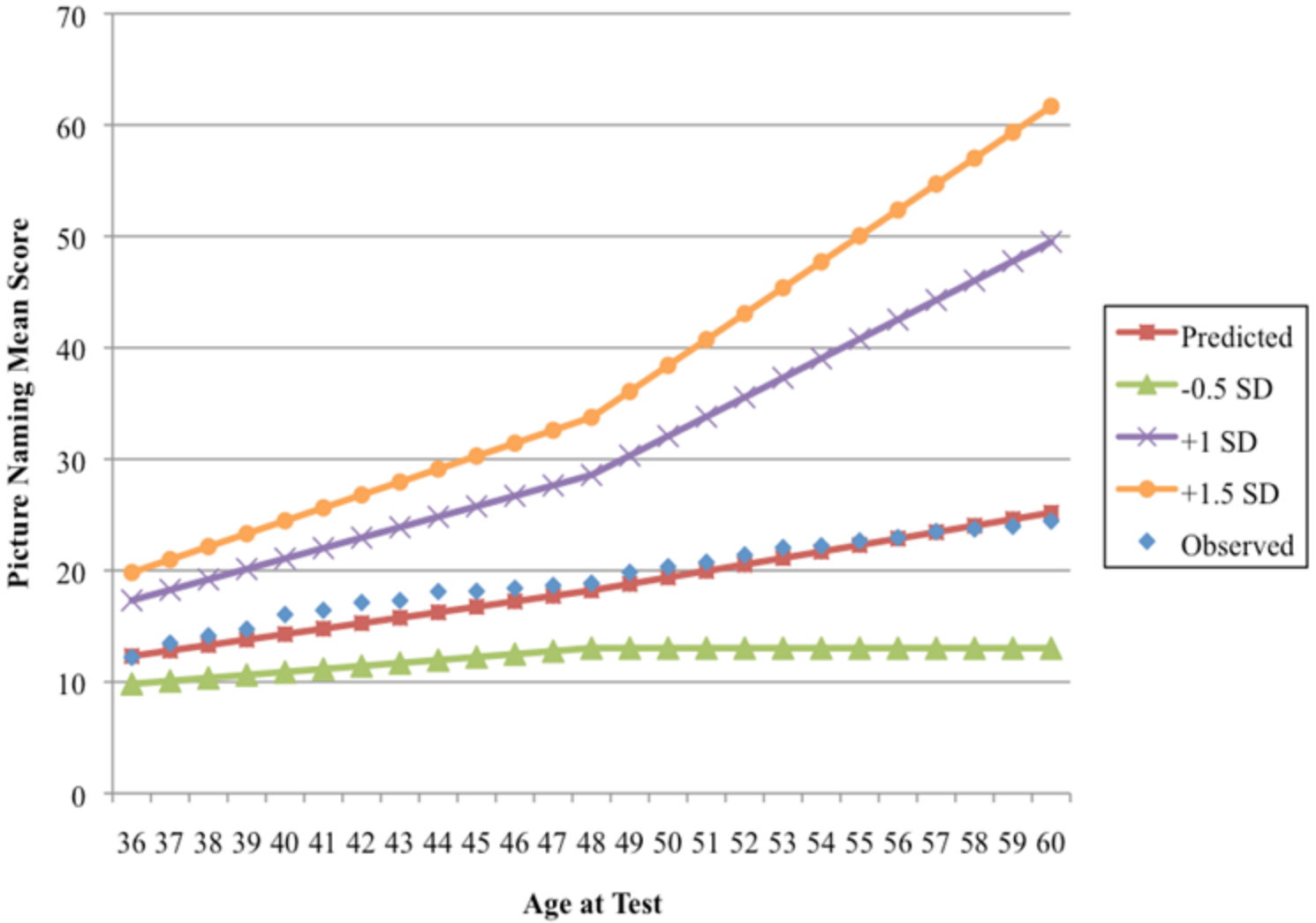
- Goal to estimate age-related growth parameters for PN, RH, and AL to fit best model for setting norm values
- 2 small, tightly-controlled studies using IGDIs with typical preschoolers informed analyses and interpretation
  - 5-month study of repeated assessment with 12 preschoolers between 44 and 68 months of age (Missall, McConnell, & Cadigan, 2006)
  - Longitudinal study of 143 children with tri-annual assessment in preschool through kindergarten (Missall et al., 2007)
- Earlier work served as guideposts with regard to considering two types of Linear Mixed Models
  - Examined a linear growth models assuming one intercept and uniform growth
  - Examined linear spline models testing growth rates between 3 and 4 year-old children
    - Knot set at 48 months based on theoretical model that K-2 growth might differ from K-1 growth, and evidence of floor effects with samples of 3's

Results: EL-IGDI Linear Growth Parameters

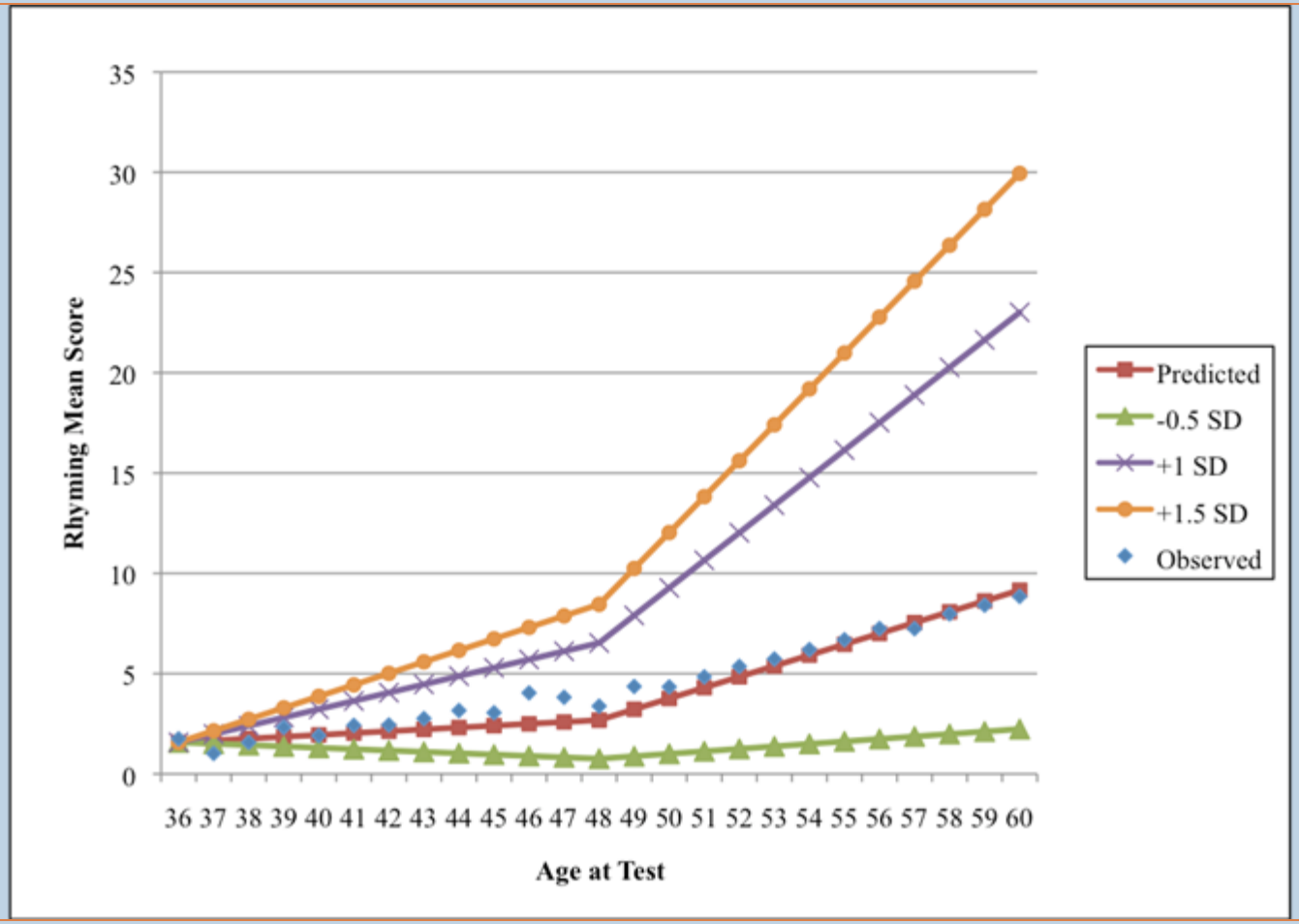
Model	Parameter/Fit	PN	RH	AL
Linear	β1	11.86	-0.55	-0.19[c]
	β2[a]	0.56	0.37	0.23
	-2LL	83,965	63,570	55,664
Spline	β1	12.59	1.50	1.42
3-year-olds	β2[b]	0.47	0.09	0.02[c]
	β3	0.12	0.47	0.32
4-year-olds	β2+β3	0.59	0.56	0.35
	-2LL	83,874	63,312	55,441

	Picture Naming			Rhyming			Alliteration		
Age/Study	2006[a]	2007[b]	U.S. Norms[c]	2006[a]	2007[b]	U.S. Norms[c]	2006[a]	2007[b]	U.S. Norms[c]
36	*7.02*	*7.48*	12.33	*< 0*	*< 0*	1.57	*4.17*	*< 0*	1.49
37	*7.68*	*8.30*	12.82	*< 0*	*< 0*	1.67	*4.38*	*< 0*	1.51
38	*8.34*	*9.11*	13.31	*< 0*	*< 0*	1.76	*4.59*	*< 0*	1.53
39	*9.00*	*9.93*	13.80	*< 0*	*< 0*	1.85	*4.80*	*< 0*	1.56
40	*9.66*	*10.75*	14.29	*< 0*	*< 0*	1.94	*5.01*	*< 0*	1.58
41	*10.32*	*11.57*	14.78	*< 0*	*0.00*	2.03	*5.22*	*< 0*	1.60
42	*10.98*	*12.39*	15.27	*< 0*	*0.50*	2.13	*5.43*	*< 0*	1.63
43	*11.64*	*13.20*	15.76	*< 0*	*1.00*	2.22	*5.64*	*< 0*	1.65
44	12.30	*14.02*	16.25	*< 0*	*1.50*	2.31	5.85	*0.20*	1.67
45	12.96	*14.84*	16.74	*< 0*	*2.00*	2.40	6.06	*0.54*	1.70
46	13.62	*15.66*	17.23	*< 0*	*2.50*	2.50	6.27	*0.87*	1.72
47	14.28	*16.48*	17.72	0.36	*3.00*	2.59	6.48	*1.21*	1.75
48	14.94	18.00	18.21	1.33	3.90	2.68	6.69	2.20	1.77
49	15.60	18.11	18.80	2.30	4.00	3.22	6.90	1.88	2.11
50	16.26	18.93	19.38	3.27	4.50	3.76	7.11	2.21	2.45
51	16.92	19.75	19.96	4.24	5.00	4.30	7.32	2.55	2.79
52	17.58	20.50	20.54	5.21	5.40	4.84	7.53	2.50	3.13
53	18.24	21.38	21.12	6.18	6.00	5.38	7.74	3.22	3.47
54	18.90	22.20	21.70	7.15	6.50	5.92	7.95	3.55	3.82
55	19.56	23.02	22.28	8.12	7.00	6.46	8.16	3.89	4.16
56	20.22	21.70	22.86	9.09	6.50	7.00	8.37	3.00	4.50
57	20.88	24.66	23.45	10.06	8.00	7.54	8.58	4.56	4.84
58	21.54	25.47	24.03	11.03	8.50	8.08	8.79	4.89	5.18
59	22.20	26.29	24.61	12.00	9.00	8.62	9.00	5.23	5.52
60	22.86	28.50	25.19	12.97	10.20	9.16	9.21	6.50	5.86

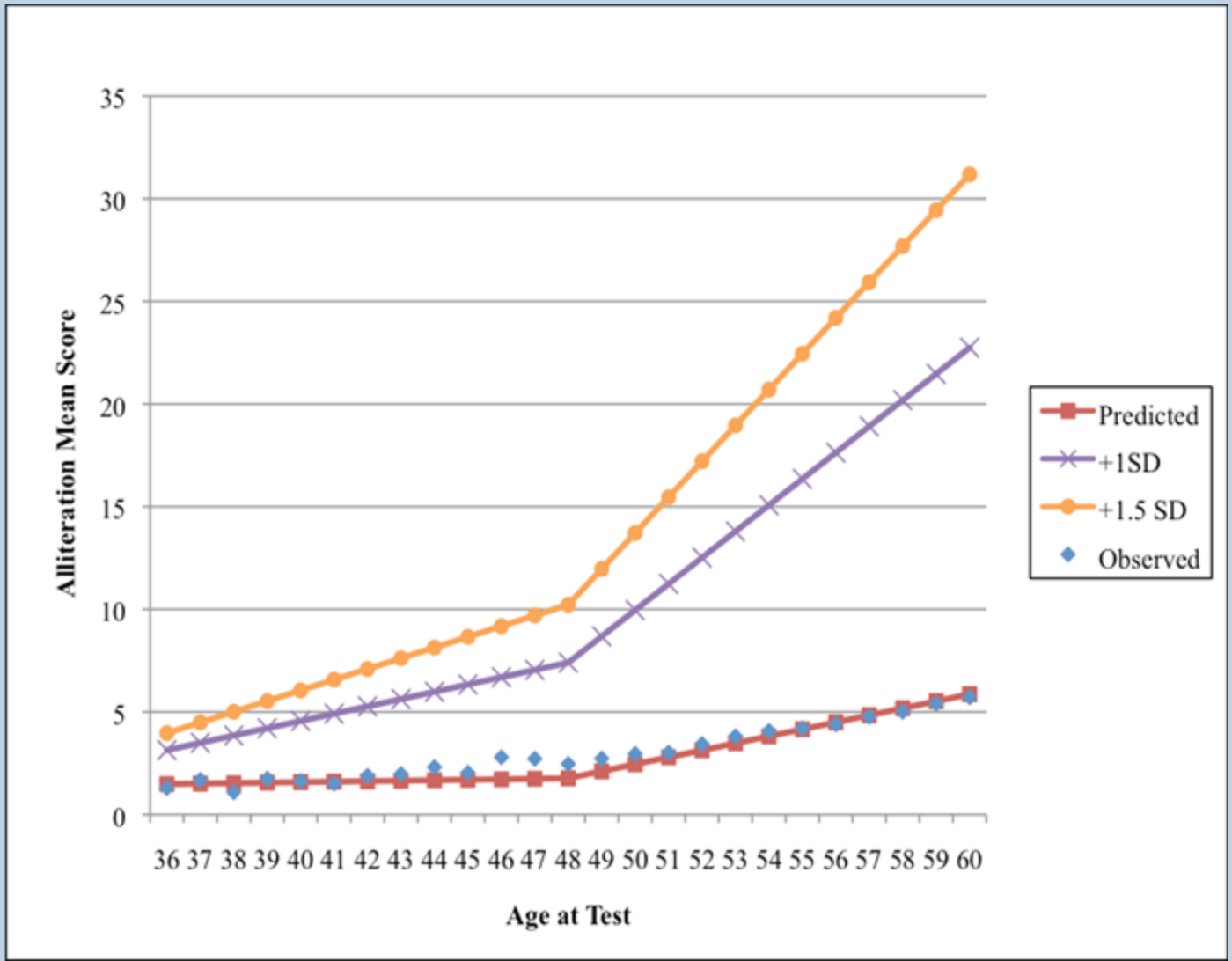
Picture Naming: Spline Model Predicted Means and Standard Deviation Intervals by Age



Rhyming: Spline Model Predicted Means and Standard Deviation Intervals by Age



Alliteration: Spline Model Predicted Means and Standard Deviation Intervals by Age



Promises

- Could this model inform similar future efforts?
  - Relative ease of data collection
  - Potential to access multiple, naturally-occurring, scale-up initiatives, and large net samples, for a fraction of the cost of a controlled study
  - Confirmation of smaller-scale studies and testing of assumptions (e.g., linear growth models)
- Capacity to influence practice relatively quickly and inform future research as efficiently
  - Professionals know when to expect children to be earning a score greater than zero and when to expect incremental gains
  - Norms clarify interpretation of growth for use of measures in progress monitoring and intervention models

Challenges/Limitations

- Analysis restricted to entered G3 variables
  - Sometimes results in assumptions
- Questions about internal validity
  - Variations in administration may or may not wash out with sample size and representation
    - Controlled here by elimination of outliers, but no control over training, administration, scoring, or data entry
  - In our case, results correspond closely with those from tightly controlled studies
    - Can be interpreted in 1 of 2 ways
      - 1) Either the measures are robust to noise….whatever variation exists did not effect results
      - 2) Measures are so easy to administer that relatively little variation exists
    - Excellent news for "RtI measures" which were designed with these features and increasingly available to interface with "field research" formats
    - Seems as though demonstrated internal validity should maximize perceptions of external/ecological validity
- Questions of representation/sampling frame/external validity
  - Selection bias in terms of who accesses and uses a system like G3?
  - How might children, preschool quality vary for users vs. non-users?
  - Even though sample similar to Census, items may function differently across gender or race/ethnicity

EMERGING PSYCHOMETRIC METHODS AND GENERAL OUTCOME MEASURES

Michael Rodriguez

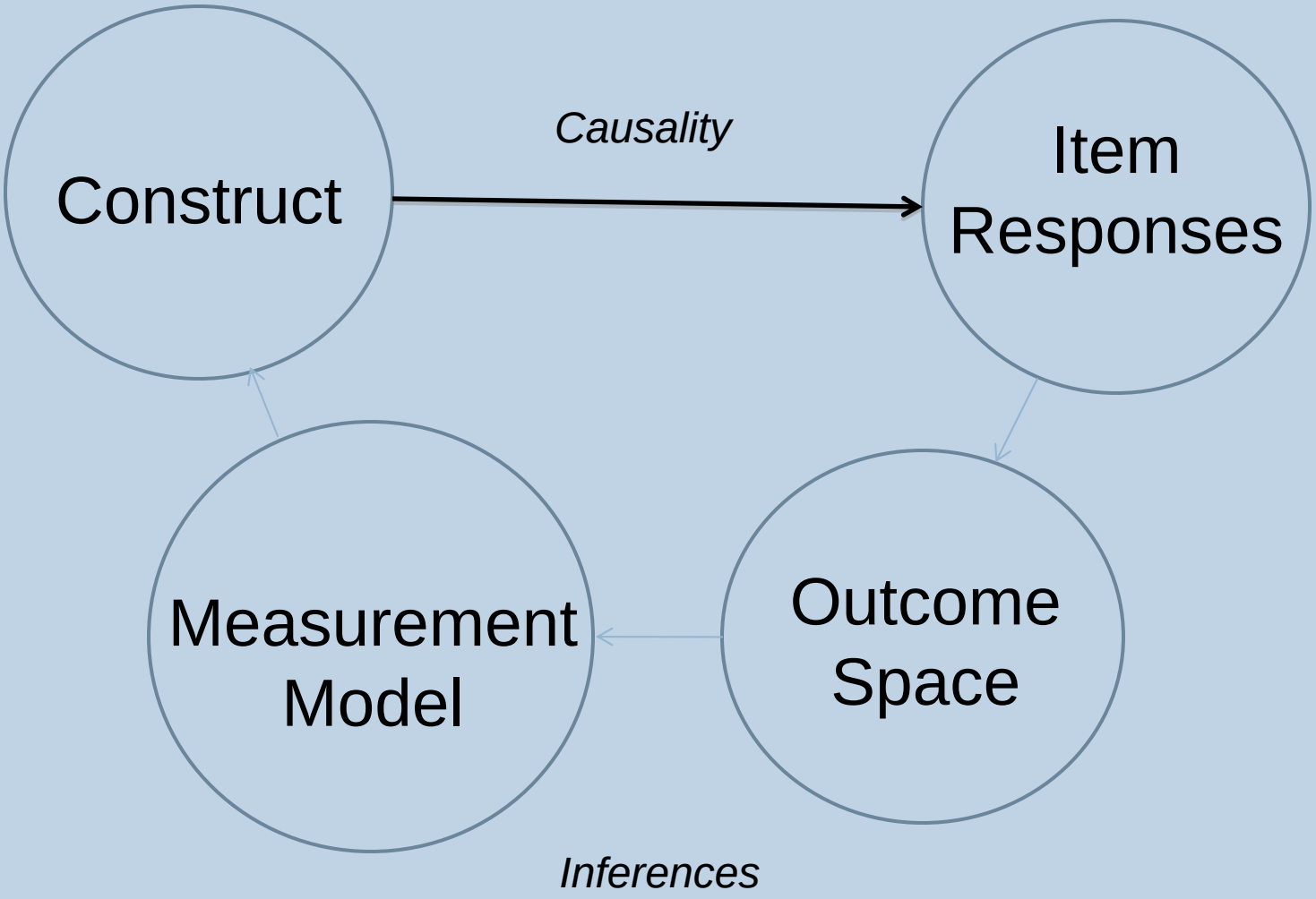
PSYCHOMETRICS TO SUPPORT RTI ASSESSMENT DESIGN

Michael C. Rodriguez
University of Minnesota

The Rasch Approach

- 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

From Construct to Item Responses



Source: Mark Wilson, 2005

Measurement Models: IRT v. Rasch

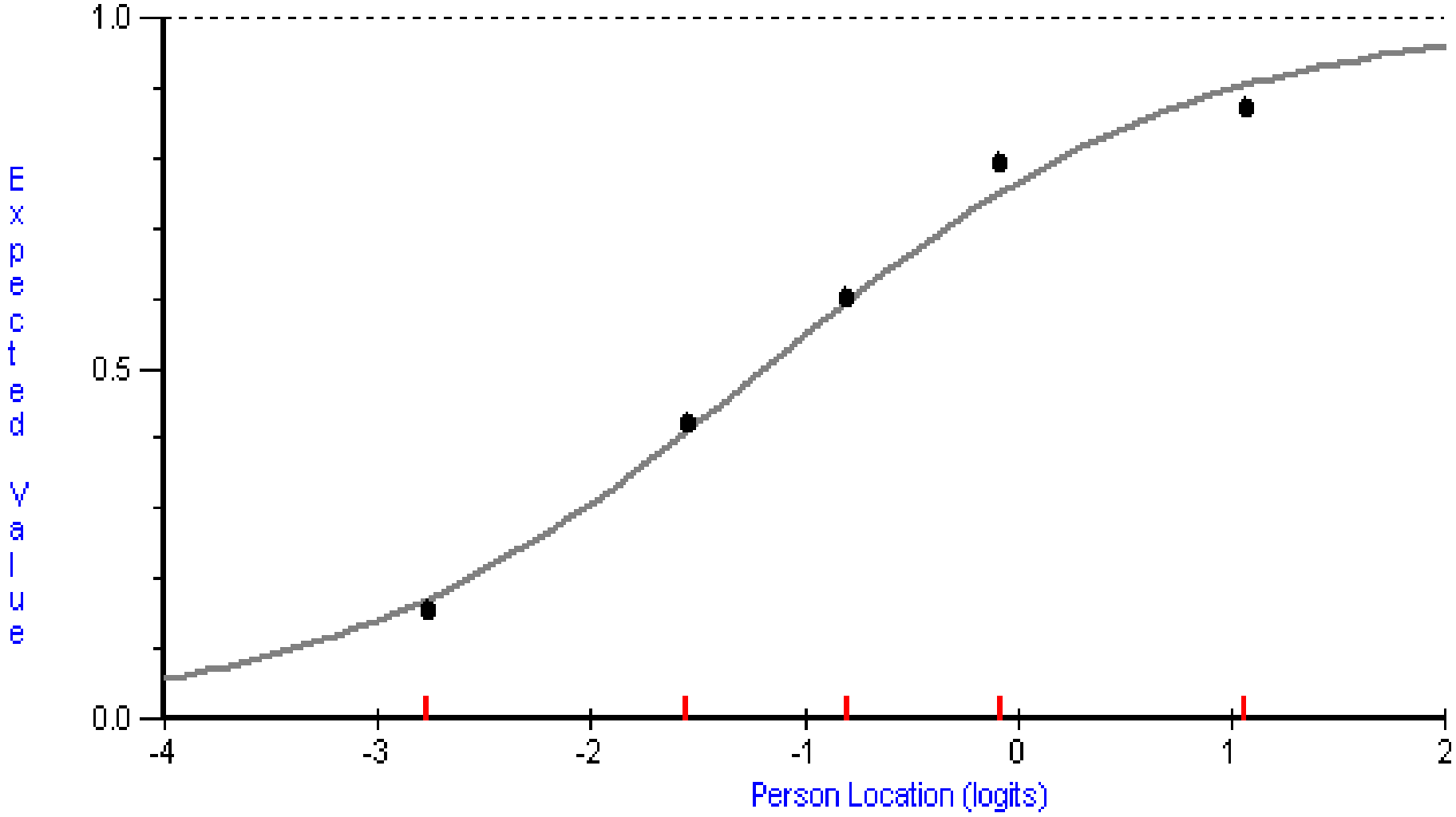
- Most IRT models are based on a paradigm that identifies a model which explains variation in the data – to find a model that best characterizes the data
- Rasch is an approach that is based on the paradigm of constructing a measure which can characterize a construct on a linear scale – such that the total score fully characterizes a person on a given construct

Rasch Philosophy

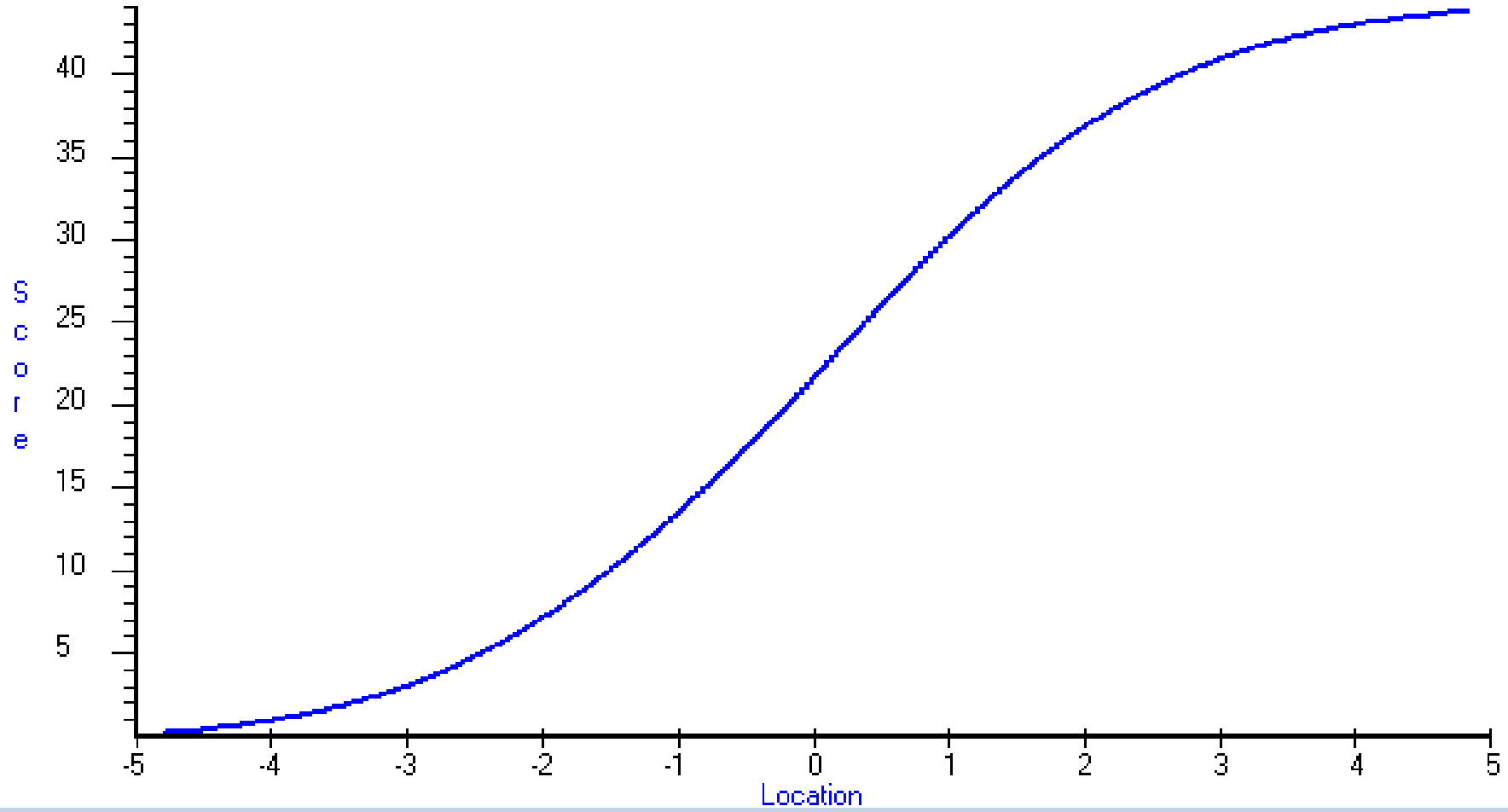
- Rasch models provide a basis and justification for obtaining person locations on a continuum from total scores on assessments.
- Although it is not uncommon to treat total scores directly as measurements, they are actually counts of discrete observations rather than measurements.

- Each observation represents the observable outcome of a **comparison** between a person and item.
- Such outcomes are directly analogous to the observation of the rotation of a balance scale in one direction or another.
- This observation would indicate that one or other object has a greater mass, but counts of such observations cannot be treated directly as measurements.

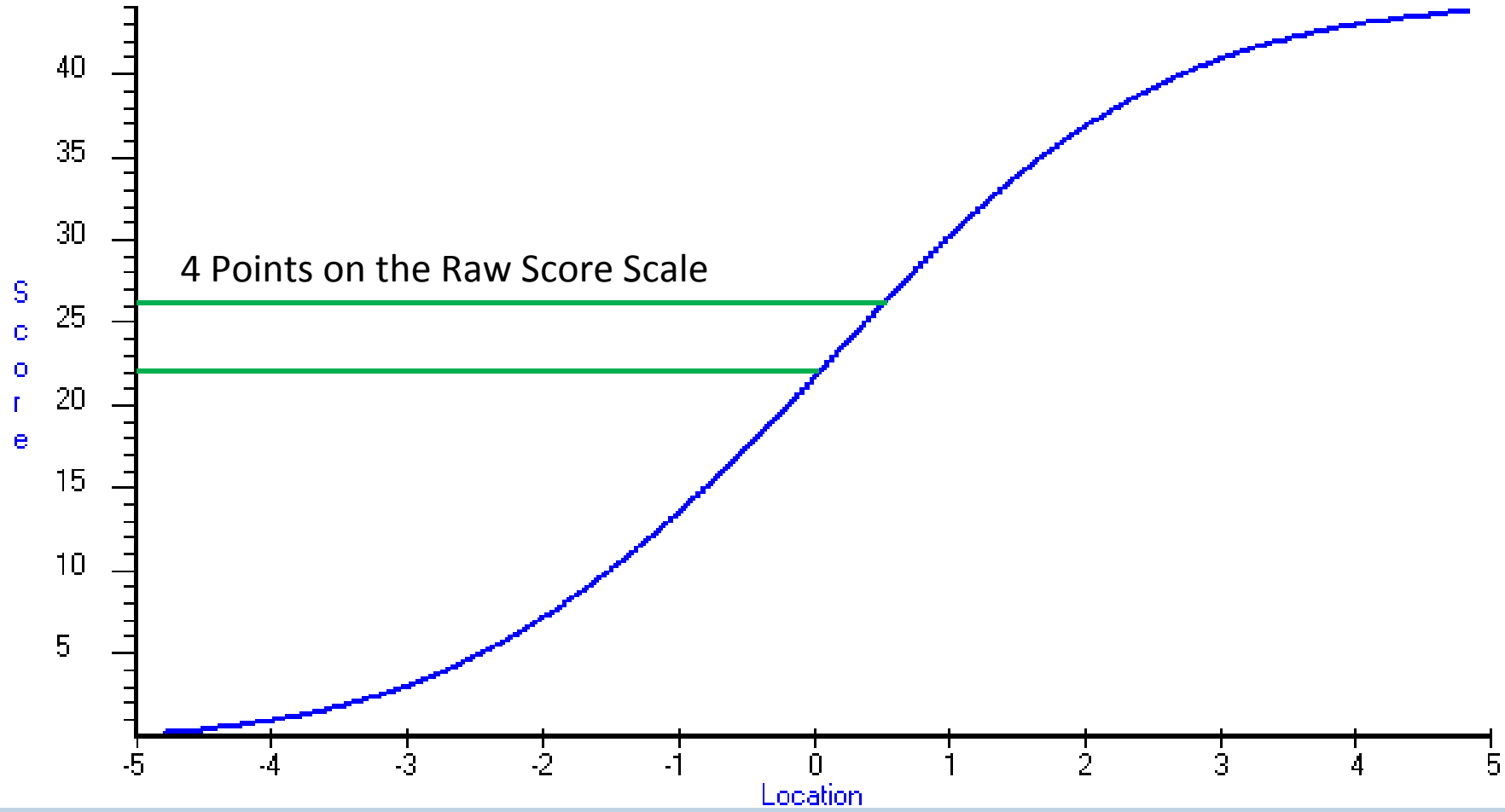
Item Characteristic Curve



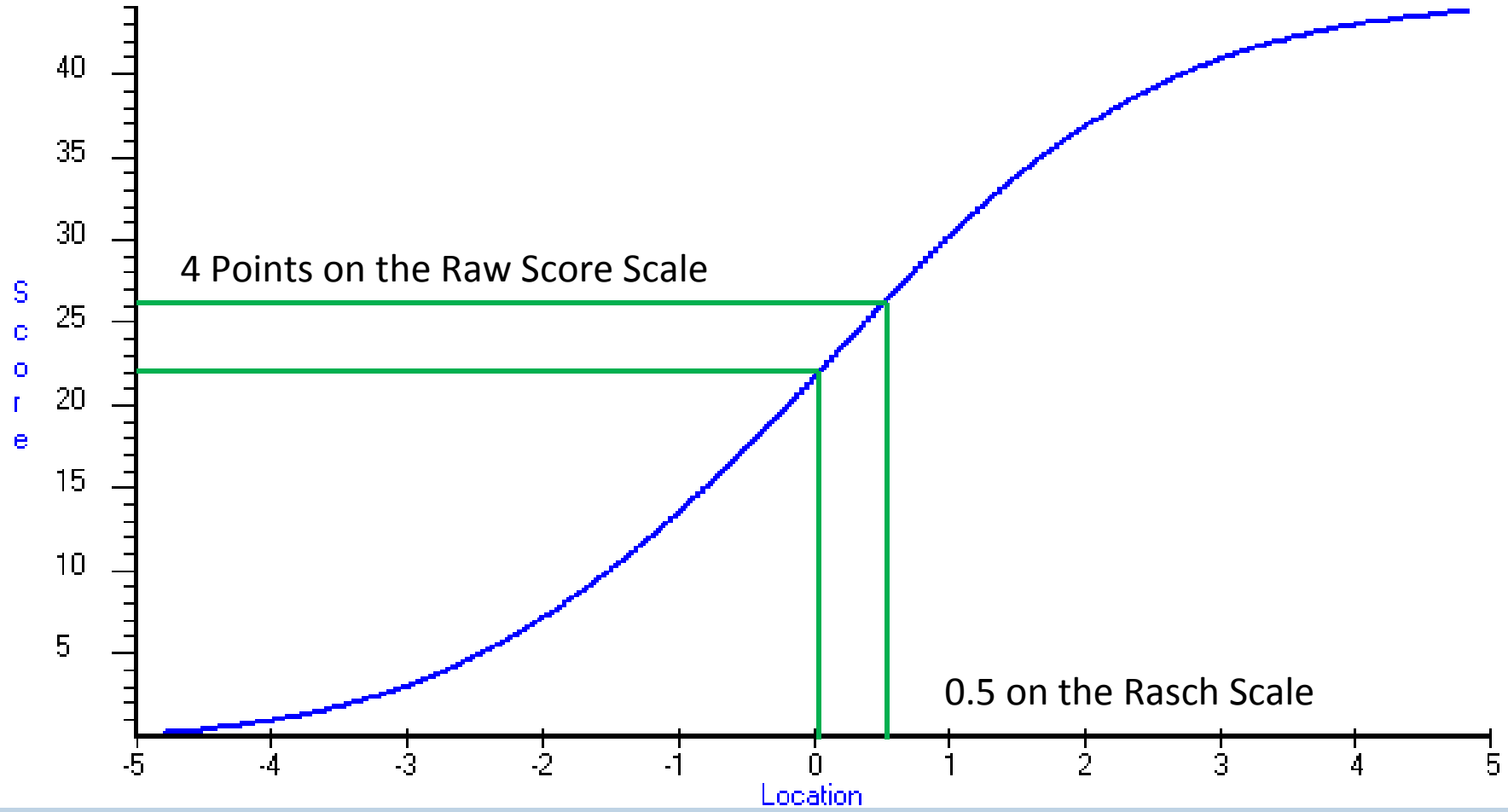
Test Characteristic Curve



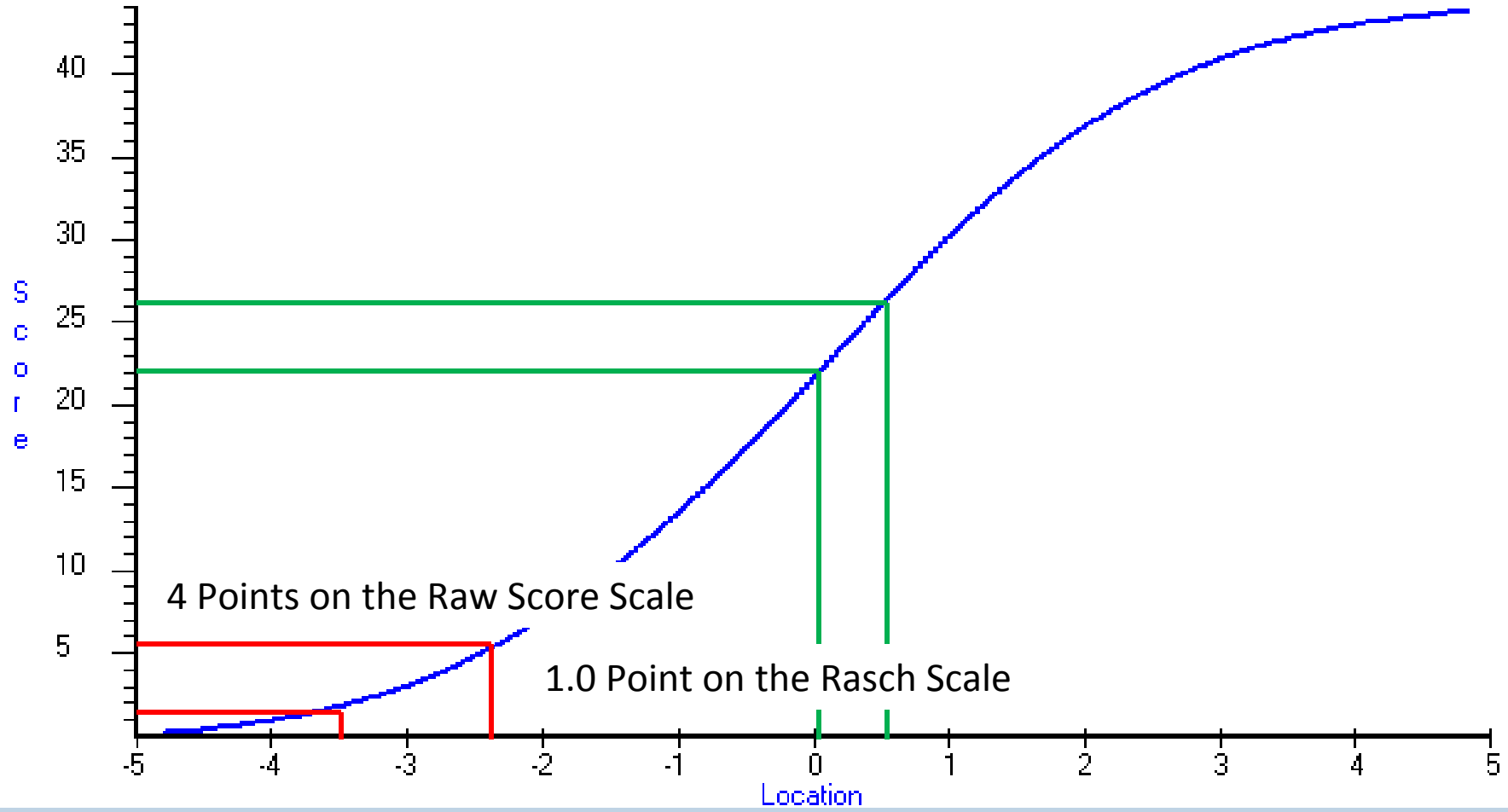
Test Characteristic Curve



Test Characteristic Curve



Test Characteristic Curve



From Numbers to Meaning

- Numbers themselves do not mean much.
  - Is 10 meters a short distance? Long distance?

We need context to bring meaning to the measure: 10 meters.

However, 10 meters should always be 10 meters, no matter who takes the measure or how it is taken.

Sample Dependent Statistics

- Is an item with a p-value of .90 easy or difficult?

  … 90% passed the item

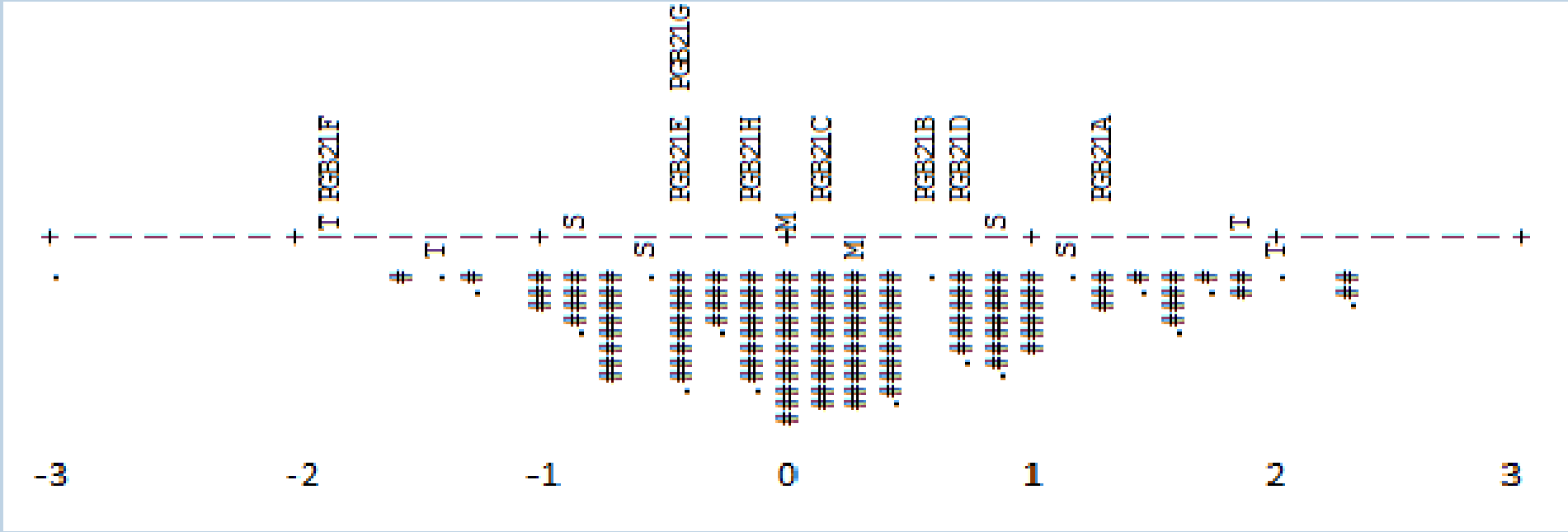
- Is a person with a score of 5 out of 50 items low in ability?

  … correctly answered 10% of the items

Rasch Scaling

- Person-free item difficult
  - Locates the items on the ability continuum
- Item-free person ability
  - Locates the person on the ability continuum
- Places items and persons on the same scale – the ITEM MAP

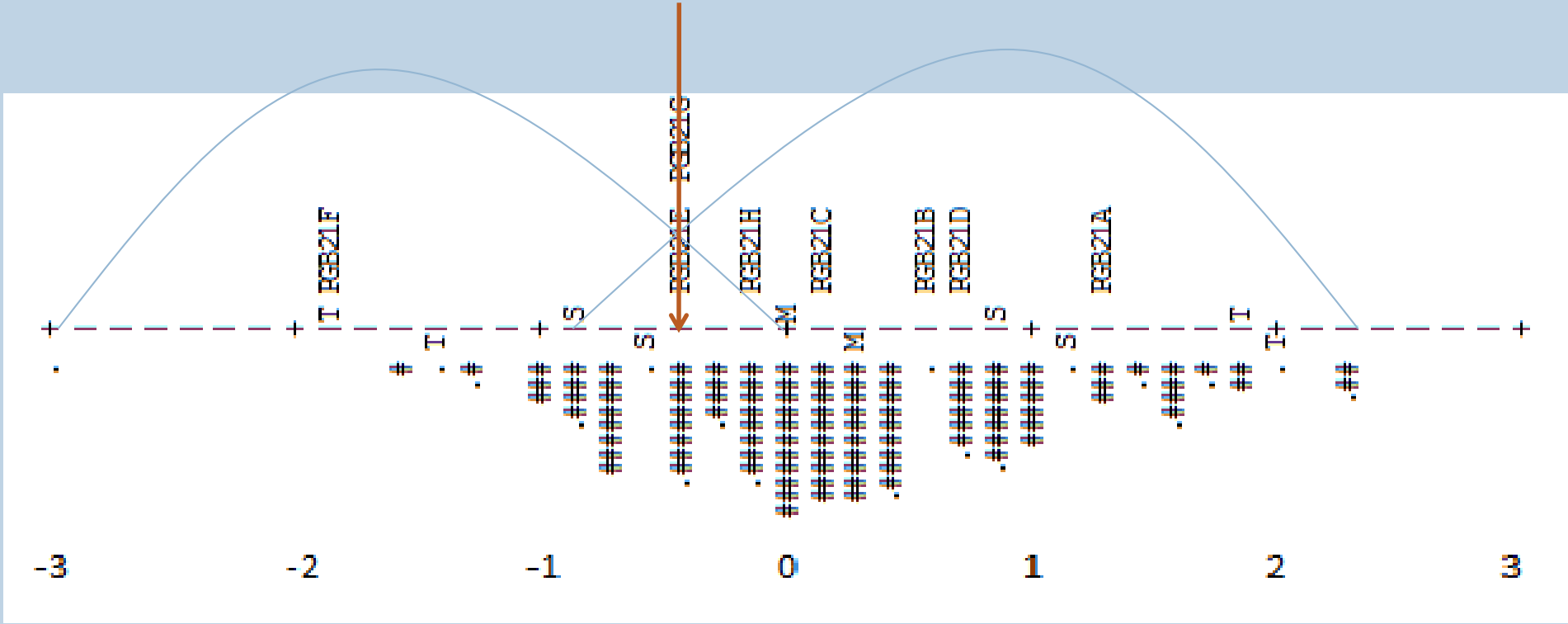
Item Map


PGB21F
PGB21G
PGB21E
PGB21H
PGB21C
PGB21B
PGB21D
PGB21A
-3
-2
-1
0
1
2
3

Construct MAP

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

Construct Map
Describing Task Characteristics



Immediate Benefits to Support IGDIs

- From limitations of relying on card difficulty
  - To locating cards along the trait continuum
- From limitations of card discrimination
  - To examining card-trait level correlations
- From ad-hoc decision making
  - To coherence in item design, item selection, scoring, analysis, interpretation, and decision making
- From a norm-referenced interpretation
  - To a criterion-referenced framework

DEVELOPING IGDIS-R

CONSIDERATIONS
CHALLENGES
& GOALS

Alisha Wackerle-Hollman
wacke020@umn.edu

Creating the IGDIs (1996-1998)

- Technical report considerations and lessons learned.
  - "IGDIs describe children's growth and development over time to indicate when performance is on-track toward some desired outcome or when different or more intensive intervention is needed (Tech Report # 4, 1998)."
    - Consider growth within a framework of development.
    - Consider service delivery (location, allocation)
    - Consider the test apparatus and environment
  - Individual items representative of the 'general outcome' desired deserve particular attention- empirically, theoretically, and culturally.

Creating the IGDIs

- General Outcome Measure (GOM) philosophy
  - Appropriate translation from reading research to early childhood (McConnell, McEvoy & Priest, 2002)
  - "General", relating to global outcomes used to predict broad-based performance in a particular content area (early literacy).

Principles of General Outcome Measures

- Ease of Use
  - Simple to administer, interpret and describe to facilitate widespread adoption with fidelity and appropriate use.
- Direct Assessment of Growth
  - Used by teachers and educational professionals to evaluate instructional programs in a quick and easy manner
- Are adaptable across children, programs and purposes
  - Point in time assessment (screening) or growth over time assessment (progress monitoring).
- Empirically Supported
  - Valid, reliable and sensitive to growth.

McConnell, McEvoy & Priest, 2002; Fuchs & Deno, 1991

Additional Principles of GOMs Noted in Research

- Lack of Instructional Hierarchies
  - GOMs do not require teachers to specify instructional hierarchies before measurement can occur, focusing on the broader, final task.
- Lack of prescriptive, compartmentalized sequences of instruction.
  - GOMs avoid decomposing curricula into teachable "pieces".
  - GOMs do not determine instructional content and procedures.

(Fuchs & Deno,1991)

Our Approach

- A Marriage of GOM and Rasch Analyses to produce:
  - IGDIs that remain true to GOM characteristics, *but* are significantly improved in the areas of:
    - *Item characteristics*
    - *Statistical strength (reliability, discrimination)*
    - *Validity*
  - Potentially, a set(s) of "located" items, particular order (s) of administration and weighted scoring procedures.

Creating IGDIs-R

- Goals

To create a valid, reliable and efficient assessment model to be used to identify and assess preschool-age children (4 years old) who:

- Are in need of additional intervention (basic screening).
- Within a Response to Intervention framework, are candidates for Tier 2 and Tier 3 interventions (sensitivity within screening).
- Within intervention, have a need for a sensitive assessment tool that can evaluate progress over brief periods of time (progress monitoring).

Creating IGDIs-R

- Challenges
  - Specificity of Rasch Model suggesting an order of items
  - Feasibility of Administration
  - Maintenance of time variable
  - Relevance of an Rtl framework
- Consideration
  - Commitment to GOM characteristics
  - Balance and marriage of GOM characteristics and Rasch Model features.

Assumptions

- We will define each domain as a "developmental trait" as recognized in Rasch modeling- that is, as a skill set that grows over time and is related to a broad long-term outcome.
- Our work will be based on principles of *both* traditional classical test theory (CTT) as used in the original IGDIs in conjunction with GOM philosophy and Rasch Modeling, with an evaluation of these models along a continuum of collaboration to determine the best fit for the IGDI-R assessment set.

(McConnell, 2010)

IGDI-R

- Research Process
  - Foundational Literature Review
  - Operational Definitions
  - Development of Measures (15 candidate measures across four domains)
  - Pilot ($n \approx 15$)
  - Phase 1 ($n \approx 40$)
  - Phase 2 ($n \approx 1000$)
  - Phase 3 (tbd)

New Measures

Oral Language

- Which One Doesn't Belong
- Definitional Vocabulary
- Picture Naming*

Phonemic Awareness

- Rhyming*
- Alliteration*
- Sound Blending

Alphabet Knowledge

- Letter Sounds
- Letter Orientation

Phonemic Awareness

- Sequencing
- Sentence Comp.
- Picture Comp.

Phase 1 & Phase 2

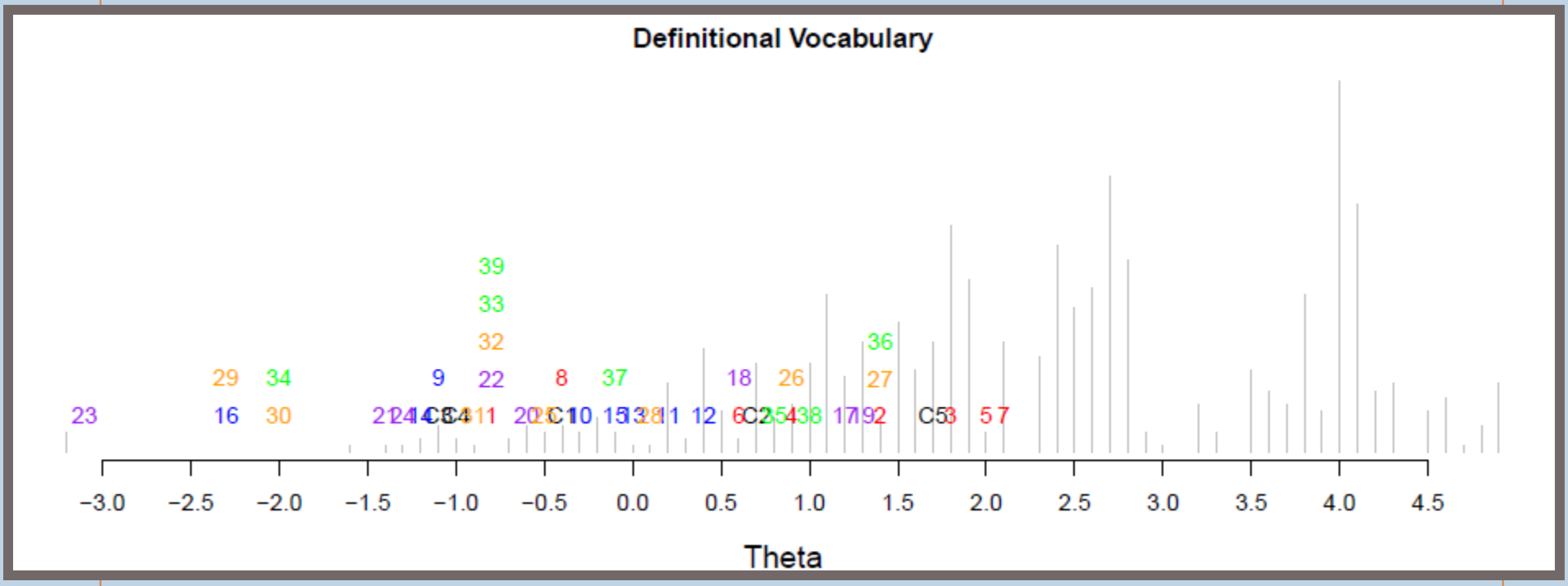
- Phase 1
  - Piloting of measures to determine feasibility with children.
- Phase 2
  - Diverse Sample across four states.
  - Examined performance:
    - At the level of the item for individual analyses.
    - As a screening tool to assess risk.
  - Changes in administration
    - Sequencing of items rather than random sample
    - Plausible Answers

Considering Selection Criteria

- Ratio Standard Deviation & Mean
  - The SD should not exceed a value of half of the mean.
- Skew & Kurtosis
  - Skew and Kurtosis should not exceed an absolute value of 1.
- Slope
  - Slope should be greater than 0 and positive.
- Percentage of zeros
  - The percentage of zero scores should not exceed 5%.

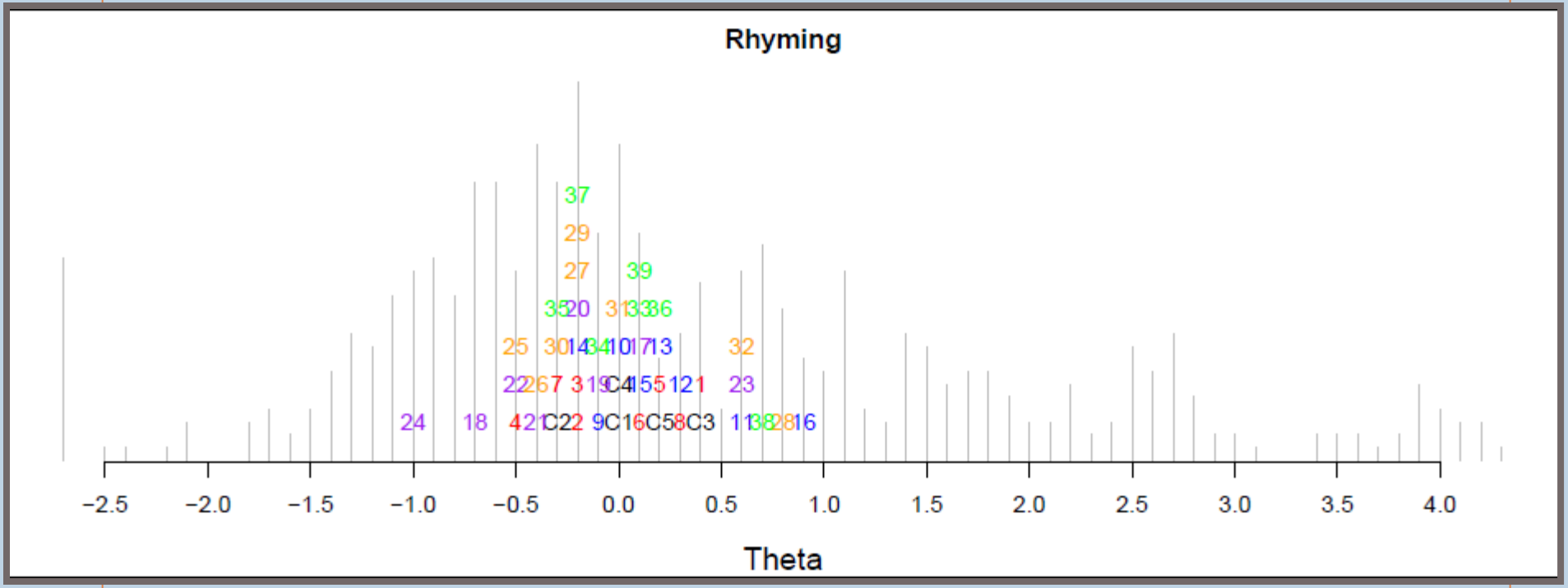
Rasch Model

- Item analysis Examples



Rasch Model

- Item analysis Examples

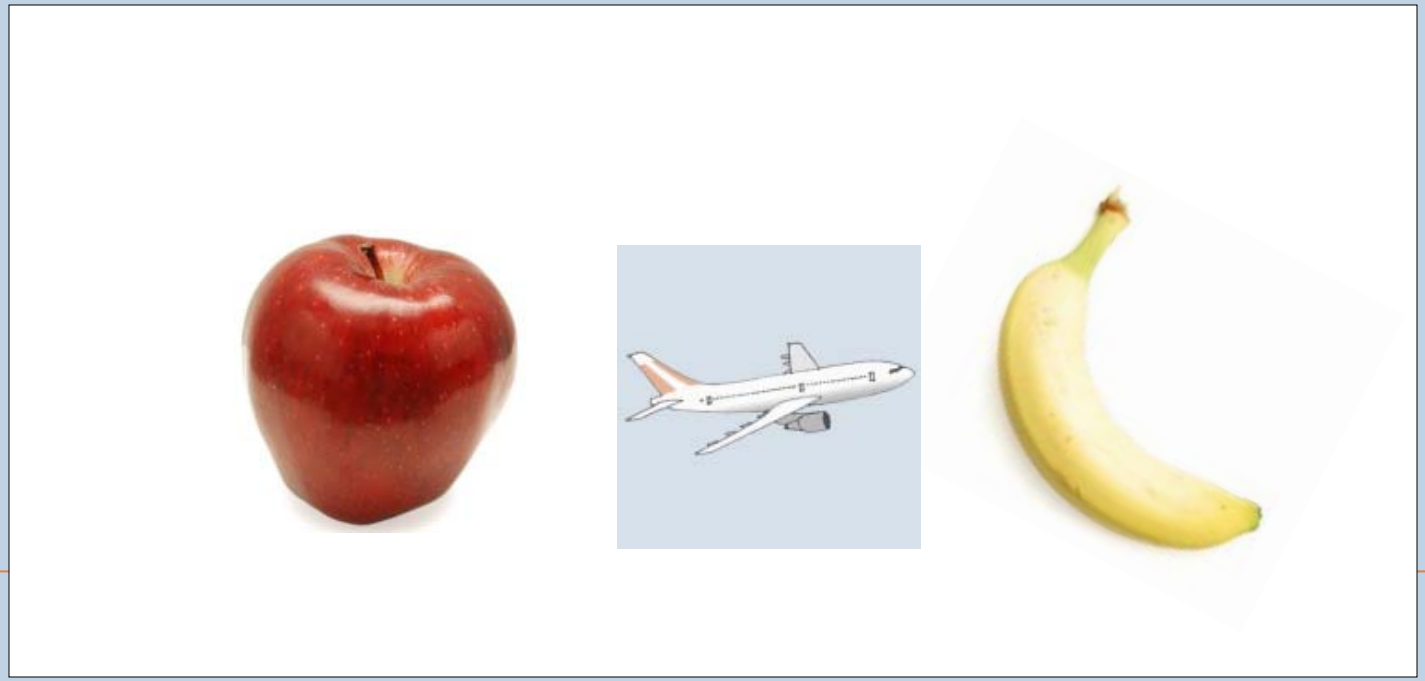


Indices of Fit

- Item Analysis
  - Difficulty Values (θ)
    - Values centered at 0 extending 2-4 standard deviations in either direction.
  - Discrimination Values
    - Offers a statistical perspective on the item's ability to discriminate between those individuals who do well on the item and those who do poorly. Discrimination is the extent to which success on an item corresponds to success on the whole test (Kelley, Ebel & Linacre, 2002).
  - Characteristics of an item's images, style, format and prompt.

Receptive Language Measures

- Which one Doesn't Belong?
  - Child points to the picture that is different from the other pictures; 2 minutes



Alliteration

Indices of Fit

- GOM characteristics
  - Maintain or Improve
    - Sensitivity
    - Adaptability
    - Robust over long-term outcomes
    - Technical Adequacy
  - Potentially Maintain
    - Assessment of Global Domain of Early Literacy
    - Efficiency
    - Cost Efficiency
  - Potentially Lose
    - Sample of skill set across the domain

Questions & Considerations

- To what extent can we efficiently and effectively marry Rasch Models with the GOM philosophy to produce meaningful data and outcomes for educators?
- What GOM characteristics remain essential when considering changes reflected in Rasch models?
- As the Rasch Models improve the technical adequacy of the IGDIs-R, what compromises must be made in terms of practical applications for educators?

After March 1st this presentation will be available at CEED's homepage: http://www.cehd.umn.edu/CEED/

ASSESSING PROGRESS IN THE DEVELOPMENT OF IGDIs AND OTHER GOMs

Scott McConnell

Two Primary Questions

- Do we have sufficient information to pose a "measurement architecture" for Response to Intervention in early childhood programs?
- What further research, and methodological innovations, are needed to meet this challenge?

Purposes of RtI Assessment System

- Identify T1 children who would benefit from T2 or T3 intervention
  - Identify domain(s) of intervention needed – oral language and vocabulary; phonological and phonemic awareness, print awareness and alphabet knowledge; comprehension
- Monitor progress for T2 and T3 students
  - Make "hold" or "move" decisions: T2 to T1, T2 to T3, T3 to T2
- Expand knowledge of GOMs and ECE–to–K-12 assessment

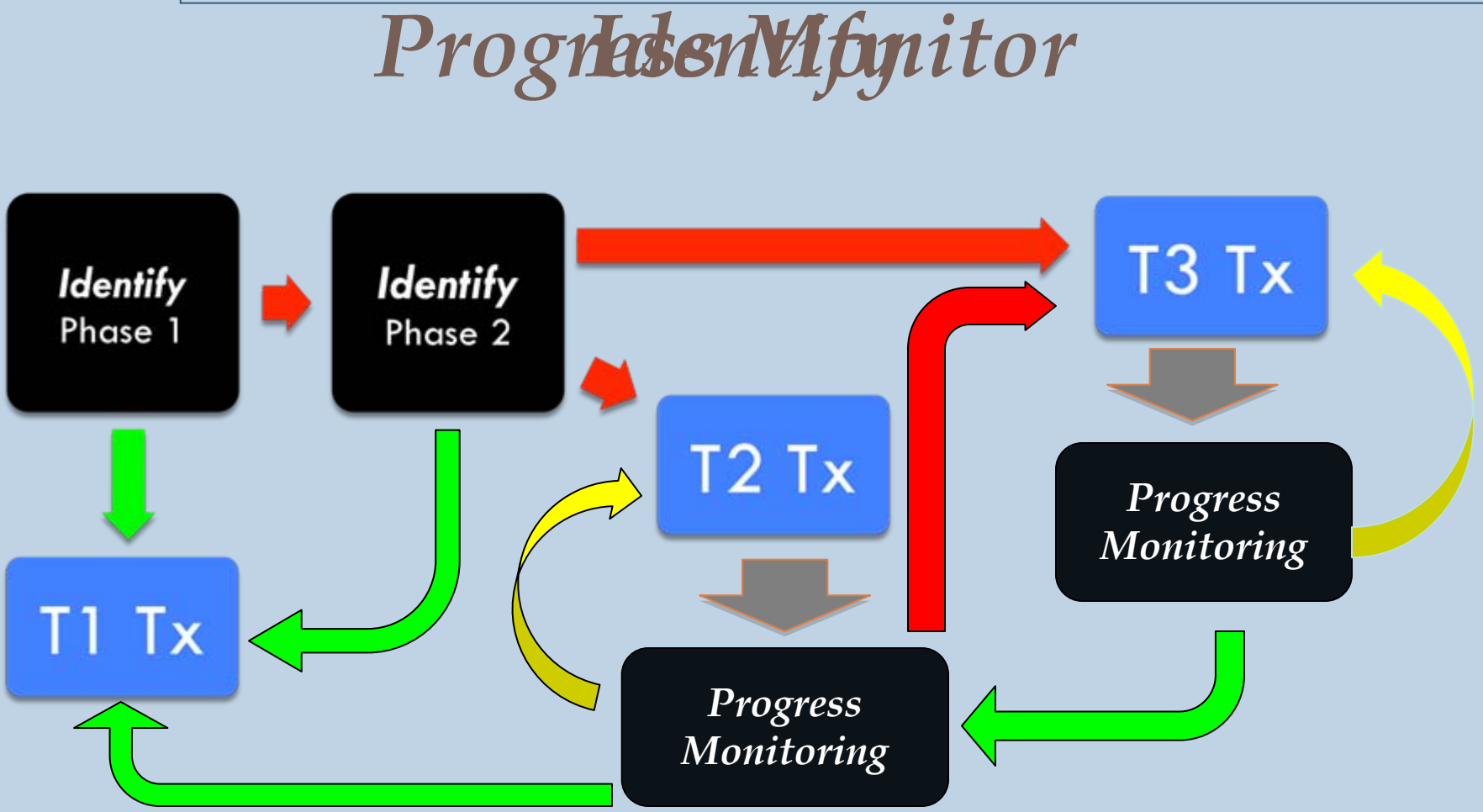
Design Principles for Measures

- GOM-ness
  - Efficient, repeatable, related to long-term outcome (what?)
  - Indicator-level assessment (not necessarily diagnostic for intervention)
- Sensitivity
  - Screening – Err in direction of false positives
  - Identification – Quarterly status
  - Progress Monitoring – Biweekly progress
- Psychometric rigor
- Feasibility

Multiple validity relations

- Concurrent validity *viz* 'criterion' measures
- Treatment validity *viz* T1, T2, and T3 interventions
- Predictive validity *viz* DIBELS, other language and literacy measures (K-3)
- Face validity
  - Teachers (but clarity about relation to treatment)
  - RtI advocates

Overview of Assessment System



"Identify"

- *Sample*
  - All enrolled children, P4
- *Frequency*
  - Quarterly (July, Sept, Dec, March)
- *Formats: Multiple-gating approach*
  - Phase 1: High-efficiency, moderate-accuracy (esp. sensitivity) screen across domains– teacher nomination?
  - Phase 2: Direct assessment, revised IGDIs by domain

"Progress Monitoring"

- Sample
  - T2 and T3 students
- Frequency
  - Weekly or bi-weekly
- Format(s)
  - IGDIs from Identify Phase 2, adapted for sensitivity
  - Curriculum-based probes by "unit"
  - Unknowns
    - Fidelity of implementation measures?
    - Diagnostic/Instructional Planning Measures?

Putting it All Together

- Quarterly assessment
  - All Children – Teacher review and nomination
  - Lowest quartile – Direct assessment in four domains, adapted/expanded IGDIs
  - Identify – Tier 1 students eligible for Tier 2 or Tier 3 intervention by domain(s) of needed intervention
- Biweekly assessment
  - Tier 2 and Tier 3 – Adapted/Expanded IGDIs
  - Move – Tier 2 or Tier 3 students to lower/higher intensity, based on growth within domain

What research is needed?

- Efficient assessment practices
  - How do we assess all students quarterly, and some kids biweekly?
- Information management burden
  - How do we collect, store, and report information in ways teachers can use it?
- Measurement infrastructure
  - What tools will be needed for "identify" and "progress monitor" assessment?
  - Will the tools be different?
- Diagnostic or functional analysis
  - What is the relation between "domain" and need for intervention?
  - How will individualized intervention be supported?

Discussion

- Samples and Norms/Benchmarks
- GOMs and IRT
- New formats, new domains, new approaches
- Where are we/Where are we going?

Discussion Questions

- This work reveals a tension between internal and external/ecological validity for "field-relevant" research…is getting stuck in traditional psychometrics too fussy and off-point?
  - Is it reasonable to assert that data from a source like G3 are procedurally corrupt? (The problem being that data from a non-controlled study can only be described in so much detail.)
- How should a sampling frame be defined? Most norm studies follow strict guidelines with regard to representation across gender, geography, race, income, education, and so on. What are the relative benefits and drawbacks to using naturally-occurring data sets?
- What types of data are going to best inform "on the floor" data-based decision making?
  - Are IGDI/CBM norms appropriate given that young children are so distinct and CBM was developed for monitoring individual growth and development?
    - What about the relative merits of local vs. national norms?
    - What about relative merit/usefulness of norms vs. benchmarks?
      - Norms: age-based expectations for a representative group
      - Benchmark: time-based expectations grounded in relation to future skills

Discussion Questions

- Samples and Norms/Benchmarks
  - What should be our comparison group(s) for setting standards?
  - What are the upsides and downsides of large samples?
- GOMs and IRT
  - To what extent can we efficiently and effectively marry Rasch Models and GOMs?
  - What GOM characteristics remain essential when using Rasch models?
  - As the Rasch Models improve the technical adequacy of the IGDIs-R, what compromises must be made in terms of practical applications for educators?
- New formats, new domains, new approaches
  - Is there a limit to what you can do with a 5.5x8.5 cards?
  - What is the relation between sensitivity, diagnostic precision, and general outcome measurement?
- Where are we/Where are we going?
  - What is possible, with precision, in various ECE settings?
  - What are the logistical possibilities and limitations for assessment, data management, and reporting/analysis?
  - How should we combine periodic screening and progress monitoring?