Rasch Measurement in Medical Rehabilitation

Allen Heinemann, Ph.D.
Northwestern University Medical School
Rehabilitation Institute of Chicago

21 August 2014 1

Webinar Objectives

- Provide an introduction to contemporary measurement approaches applicable to medical rehabilitation
- Describe rating scale (Rasch) analysis
- Illustrate applications of rating scale analysis in medical rehabilitation

21 August 2014 2

Rehabilitation Measurement Problems

- Does a total FIM™ score of 70 obtained from two different patients (e.g., SCI, CVA) imply the same level of disability?
- Is Mary's FIM™ improvement from 13 to 23 equivalent to Sam's improvement from 116 to 126?
- How do we know these items 'add up' to mean anything?
- Can we apply arithmetic to numbers like these?

More Rehabilitation Measurement Problems

- Is the improvement between a FIM™ rating scale point of 1 & 2 equivalent to the distance between 5 & 6?
- Does the Barthel Index item weighting affect conclusions about disability severity?
- Rating scale definitions are arbitrary
- Item weights reflect instrument developers' biases

Misunderstandings about Measurement

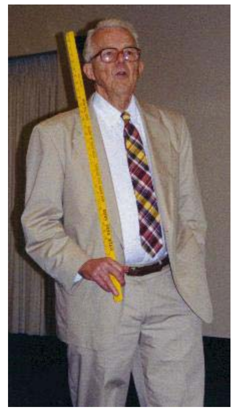
- Limit our ability to construct measures of patients' impairment, disability, handicap
- Create opportunities for "misinference"
- Make discussions of improvement difficult
- Limit our ability to compare patients' with different underlying conditions

Defining Measurement

- Determination of the magnitude of a specified attribute of an object, organism, event
- Acquired purposefully
- A fixed unit
- Expressed by a numeral

Example: Height

- Attribute: Height
- Measuring instrument: Ruler
- Magnitude of attribute: Distance from toe to head
- Fixed unit: Centimeter
- Expressed as a numeral: Height equals some number of centimeters



21 August 2014 7

The Ruler as a Measurement System

- A ruler is independent of its construction
- A ruler implements the idea of a unit
- The unit is the same everywhere along a ruler
- The centimeter is the concrete realization of the unit
- Concrete units operationalize additivity
- Rulers employ a natural number system
- Rulers implement a correspondence between units and numbers

21 August 2014 After MH Stone, RMT, 1996 8

Measurement within a Larger Perspective of Scientific Inquiry

- e**X**perience
- **O**bservation
- **M**easurement
- **A**nalysis
- **T**heory

After BD Wright

21 August 2014 9

Where Measurement Belongs in Scientific Inquiry

- e**X**perience
- **O**bservation
- **M**easurement
- **A**nalysis
- **T**heory

21 August 2014 10

How Constructs Achieve Definition

- Personal knowledge, intuition, subjectivity
  - too hot to touch –– as cold as snow
  - safe radiation exposure –– unsafe exposure
- Data-based instrument calibrations
  - test-specific functional status
- Theory-based instrument calibrations
  - manufactured thermometers
  - Lexile framework for reading comprehension

21 August 2014 After AJ Stenner & I Horabin, RMT, 1992 11

How Do Numbers Occur?

- As **labels**
  - Zip codes, house addresses, product codes
- As **counts, scores, ranks**
  - FIM^{TM} scores, Olympic placements
- As **measures**: Reliable linear units
  - Hourly pay rate, Celsius degrees

21 August 2014 12

Levels of Measurement

- Nominal (categorical)
- Ordinal (counts, ordered categories–20s, 30s)

- Equal-Interval (linear)
- Ratio (local or absolute zero defined)

Stevens, 1946, 1951

21 August 2014 13

Kinston's Levels of Representation

I. Entity (*idea*)
making a distinction (identifying, distinguishing)
II. Observable (*thingness*)
making an intersection (pointing to, counting)
III. Comparable (*quantity*)
making a comparison (ordering, ranking, valuing)
IV. Measurable (*standard unit*)
making a measurement (measuring, using arithmetic)
V. Relatable
making a relation (connecting, formulating)

W Kinston: Structure of Scientific Analysis, *Systems Research, 2,* 1985

21 August 2014 14

Kinston's Levels of Representation: A Rehabilitation Example

I. Entity (*an idea or concept*)
Disability
II. Observable (*thingness*)
Wheelchair mobility
III. Comparable (*a quantity*, *possessed by someone*)
More or less assistance, skill level
IV. Measurable (*standard unit*)
Rehabit
V. Relatable
Disability is reduced by environmental modification

21 August 2014 15

Subjectivity and Objectivity

Level	Site of Experience	Relevant Notion	Error Risk
I. Entity	Subjective	Plausibility	Losing contact with external reality
II. Observable	Objective	Probability	Making wrong distinctions, using unsuitable categories
III. Comparable	Subjective	Plausibility	Emphasizing irrelevant attributes
IV. Measurable	Objective	Probability	Examining unpromising relations, forgetting to control for crucial factors
V. Relatable	Subjective	Plausibility	Producing findings of limited applicability

21 August 2014 16

We Use Numbers to *Compare and Evaluate*

- 10 rehab days @ $1,000/day > 7 rehab days @ $1,200
- Home discharge after 25 rehab days is a better outcome than nursing home discharge after 15 days
- A 50 FIM^{TM} point gain over 10 days is an average 5 point gain/day

21 August 2014 17

This Kind of Number Use Requires Arithmetic

- Add, subtract, multiply, average
- Arithmetic done on labels, raw scores, counts and ranks is nonsense
- Conclusions can be misleading
- Equal-interval, constant unit, linear measures are required

21 August 2014 18

Our Assumptions

- We judge correctly,
- According to standard criteria,
- With ratings recorded accurately, and that
- 1,2,3,4,5,6,7 are real measures for task performance that add up to scores as good as measures

Are Unsound!

21 August 2014 19

The Reality

- Ratings are merely guesses
- Standards vary by time and person
- Rating can be recorded incorrectly
- 1,2,3,4,5,6,7 are merely ranks that may not add up
- Raw scores are never measures

21 August 2014 20

Transcending Scores: Constructing Measures

- All we can observe are counts and scores
- They may provide the basis for constructing measures
- We need a mature understanding of score meaning
- What matters is not what a score is but what it implies
- We need a method for using counts and scores to estimate the measures we want them to imply

21 August 2014 21

This Method is called

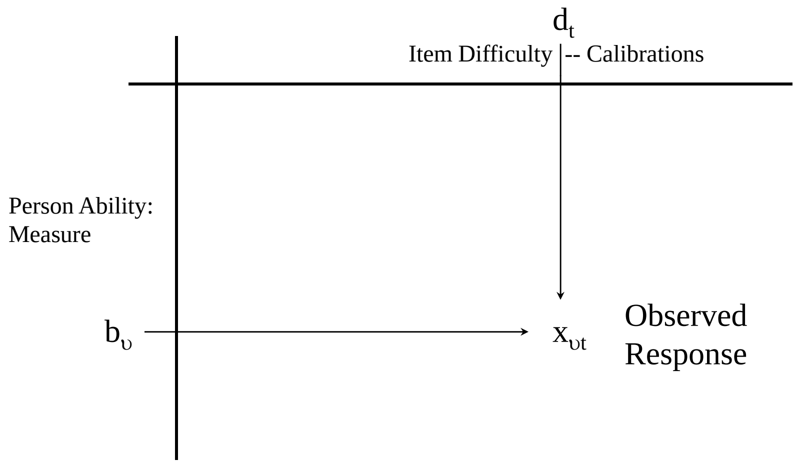
Rasch Measurement

or

Rating Scale Analysis

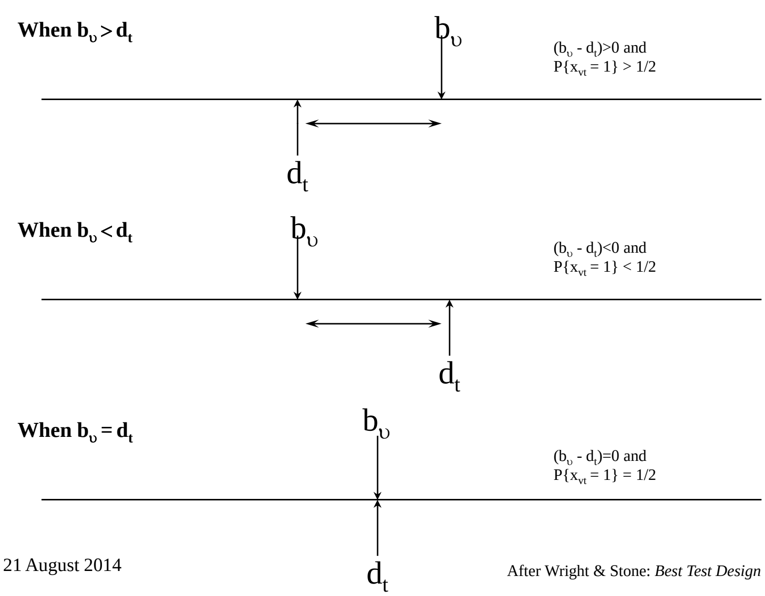
21 August 2014 22

Essential Conditions Causing a Response



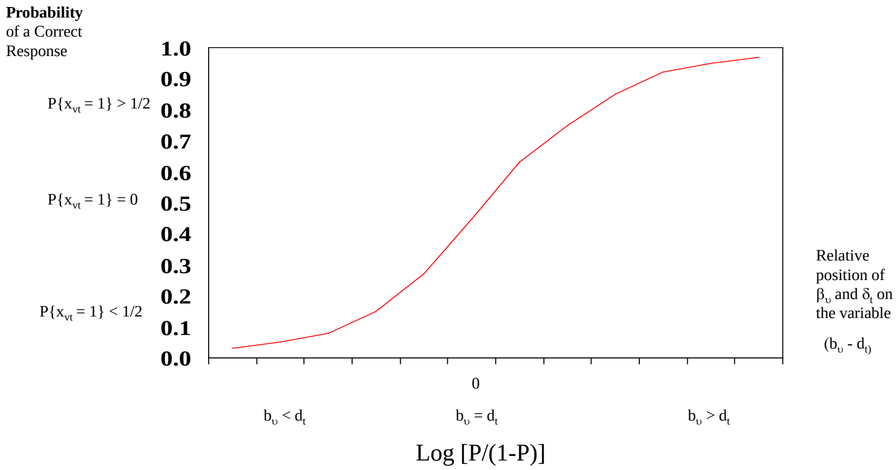
21 August 2014 After Wright & Stone: *Best Test Design* 23

How Differences between Person Ability and Item Difficulty Ought to Affect the Probability of a Correct Response



21 August 2014 After Wright & Stone: *Best Test Design* 24

The Logistic Response Curve



21 August 2014 25

Logits

- Log-odds units
- log [probability of success/probability of failure] = Person ability - Item Difficulty
- Can be estimated independently of the items included in a particular test
- More complex models can be specified

21 August 2014 26

A More Complex Example: Rating Functional Status

Observable Categories	Can't do	Needs help	Does OK
Numeric Labels	0	1	2
Interpreted to imply category probabilities	P_0	P_1	P_2
With transition odds		P_1/ P_0	P_2/ P_1

This interpretation gives the observed raw ratings inferential meaning.

21 August 2014 27

Conjoint Additivity

When these transition odds are modeled as the consequence of conjoint additivity parameters like

$$\log[P_{nijx}/P_{nijx-1}] = B_n - D_i$$

then statistical analyses of raw categorical ratings can *construct* and *control* inferentially stable linear measures, where

21 August 2014 28

The Model Parameters are defined as

- the Functional Measure $\mathbf{B_n}$ of patient n, when rated
- on Task i of Difficulty $\mathbf{D_i}$

21 August 2014 29

Additional Model Parameters are defined as

- the Functional Measure $\mathbf{B_n}$ of patient n, when rated
- on Task i of Difficulty $\mathbf{D_i}$

- by Clinician j of Rigor $\mathbf{C_j}$
- at Step x with Gradient $\mathbf{F_x}$

21 August 2014 30

Rating Scale Information

- Difficulty of each Task
- Measure for each Person
- Rigor of each Rater
- Gradient of each rating Step, ***and***

- Fit of each task to an underlying construct
- Fit of each person to this construct
- Separation of sample into distinct strata

21 August 2014 31

How Good are the Measures Produced by the Data?

- Do data fit the measurement model?
- Quality control method is needed
- Fit statistics ensure that all
  - items measure the same construct
  - people possess the same quality
  - judges understand the same rating scheme

21 August 2014 32

Fit Statistics

- Infit
  - Dominated by unexpected inlying patterns among informative, on-target observations
  - Inlier sensitive ($B_v \sim D_i$)
- Outfit
  - Dominated by unexpected outlying, off-target, low information responses
  - Outlier sensitive ($B_v < D_i$ or $B_v > D_i$)

21 August 2014 33

Fit Problems

- Muted Patterns (small fit statistics)
  - Unmodeled dependence
  - Redundancy
  - Error trends
- Noisy Patterns (large fit statistics)
  - Unexpected, unrelated irregularities

21 August 2014 34

Reliability and Separation

- Observed variance = “True” variance + error variance
- Reliability = “True” variance/Error variance
- Separation =√ [reliability / (1-reliability)]
  = “True” SD / Error SD
- Separation is the number of statistically different performance strata the test can identify in the sample

21 August 2014 35

Person Separation

- The ability of the items to resolve the sample into distinct strata
- Describes the spread of the persons *in this sample*
- “Real” separation adjusts for misfitting persons

21 August 2014 36

Relationship between Reliability and Separation

Separation Ratio G	KR-20 Alpha $G^2/(1+G^2)$	% Variance: Not Due to Error/Due to Error	Distinct Strata (4G+1)/3
0	**.00**	**0/100**	**1**
1	**.50**	**50/50**	**1**
1.5	**.70**	**70/30**	**2**
2	**.80**	**80/20**	**3**
3	**.90**	**90/10**	**4**
4	**.94**	**94/6**	**5**
5	**.96**	**96/4**	**7**
6	**.97**	**97/3**	**8**
7	**.98**	**98/2**	**9**

WP Fisher, RMT, 1992

21 August 2014 37

Common Problems with Rating Structures

- More categories provided than respondents differentiate
- Some categories vaguely defined, e.g.,
  - FIM™ assistance required
  - "often" vs. "frequently" vs. "sometimes"
- Item difficulties become inverted
- Item fit worsens

21 August 2014 38

Sample Size and Item Calibration Stability: Dichotomous Items

Item Calibrations stable within	Confidence	Minimum sample size range (best to poor targeting)	Size for most purposes
± 1 logit	95%	16-36	30
± 1 logit	99%	27-61	50
± ½ logit	95%	64-144	100
± ½ logit	99%	108-243	150

Rating scales need smaller samples than do dichotomous items.

21 August 2014 After JM Linacre: RMT, 1994 39

How is Factor Analysis Different from Rasch Analysis?

- Same kind of data, but
  - different interpretations of numerical status
- Same estimation methods, but
  - different measurement models
- Solve the same problems, but
  - substantially different utility

21 August 2014 40

Problems with Factor Analysis

- Factor scores are not linear measures
- Factor structure is sample dependent
- Factor structure reflects rating scale used
- Ordinal variables and highly correlated factors create confusion
- Complete data are required
- Identity of smaller factors is overwhelmed by residuals from preceding larger factors
- Guidelines for stopping are vague
- Standard errors for factor scores hard to obtain

21 August 2014 After BD Wright, RMT, 1994 41

In summary: Rating scale analysis provides

- Person and Item Separation
- Person and Item Fit Statistics
- Person Ability and Item Difficulty Estimates
- Rating Scale Adequacy Indices

21 August 2014 42

Suggested Procedure for Improving Rating Scales

- Maximize person separation (2.0 minimum)
- Assure that item step structures are clear
- Examine item fit statistics
- Identify misfitting persons
- Examine item targeting on sample
- Review person and item map to establish construct

21 August 2014 43

Strategies for Improving Rating Scales

- Consider single vs. multiple rating scale model
- Verify rating scale behavior
- Survey the range of item difficulties, fit statistics
- Consider range of person ability estimates, fit statistics, subgroup possibilities

21 August 2014 44

Answers to Rehabilitation Measurement Problems

- Does a total FIM™ score of 70 obtained from two different patients (e.g., SCI, CVA) imply the same level of disability?
- Is Mary's FIM™ improvement from 13 to 23 equivalent to Sam's improvement from 116 to 126?

- Approximately the same; disability measure can be fine-tuned for specific impairments
- No - raw scores aren't measures

21 August 2014 45

Answers to More Rehabilitation Measurement Problems

- Is the improvement between a FIM™ rating scale point of 1 & 2 equivalent to the distance between 5 & 6?
- Does the Barthel Index item weighting affect conclusions about disability severity?

- No. The improvement can be calibrated
- Yes, it confounds measurement

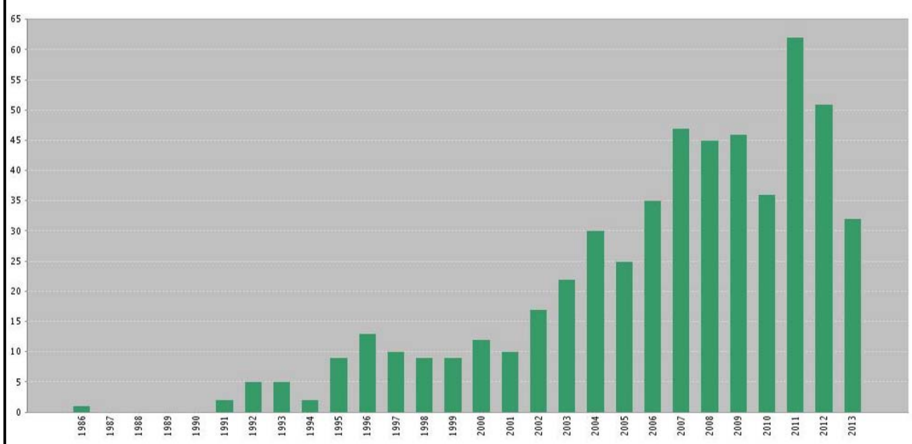
21 August 2014 46

Examples of Rasch-Developed Measures

21 August 2014 47

Rasch articles 1986-August 2013

(Rasch + Rehabilitation, Web of Knowledge)



48

Where were these 525 articles published?

- Archives of Physical Medicine and Rehabilitation
- Journal of Rehabilitation Medicine
- Disability and Rehabilitation
- Investigative Ophthalmology Visual Science

49

November 2013 APMR articles with "Rasch" keyword

- Hays: Upper-Extremity and Mobility Subdomains From the Patient-Reported Outcomes Measurement Information System (PROMIS) Adult Physical Functioning Item Bank
- Heinemann: Measuring Enfranchisement: Importance of and Control Over Participation by People With Disabilities
- Angst: Effects of Inpatient Rehabilitation in Hip and Knee Osteoarthritis: A Naturalistic Prospective Cohort Study With Intraindividual Control of Effects
- Major: Validity and Reliability of the Berg Balance Scale for Community-Dwelling Persons With Lower-Limb Amputation

50

MetaMetrics – Lexile Framework

TEXT LEXILE RANGES TO GUIDE READING FOR COLLEGE AND CAREER READINESS

GRADES	CCSS LEXILE TEXT RANGE
11–12	1185L–1385L
9–10	1050L–1335L
6–8	925L–1185L
4–5	740L–1010L
2–3	420L–820L
1	190L–530L

21 August 2014 51

THE LEXILE® FRAMEWORK FOR READING MAP

Matching Readers with Text

SAMPLE TITLES

1500L+

LITERATURE TITLES
- 1640L The Plot Against America (ROTH)
- 1560L Rob Roy (SCOTT)
- 1530L The Good Earth (BUCK)
- 1520L A Fable (FAULKNER)
- 1500L The Decameron (BOCCACCIO)

INFORMATIONAL TITLES
- 1600L Sustaining Life: How Human Health Depends on Biodiversity (CHIVIAN & BERNSTEIN)
- 1550L The Art of War (SUN TZU)
- 1520L Fair Play: The Ethics of Sport (SIMON)
- 1500L Critique of Pure Reason (KANT)

1400L

- 1460L The Legend of Sleepy Hollow (IRVING)
- 1450L Billy Budd** (MELVILLE)
- 1430L The Story of King Arthur and His Knights (PYLE)
- 1420L Life All Around Me by Ellen Foster (GIBBONS)
- 1420L The Scarlet Letter** (HAWTHORNE)

- 1490L America's Constitution: A Biography** (AMAR)
- 1490L Gettysburg Address (LINCOLN)
- 1480L The Declaration of Independence
- 1410L Profiles in Courage (KENNEDY)
- 1400L The Life and Times of Frederick Douglass (DOUGLASS)

1300L

- 1360L Robinson Crusoe (DEFOE)
- 1350L The Secret Sharer (CONRAD)
- 1340L The Hunchback of Notre Dame (HUGO)
- 1340L The Metamorphosis** (KAFKA)
- 1340L Fever Pitch (HORNBY)

- 1390L In Defense of Food: An Eater's Manifesto (POLLAN)
- 1370L Jane Austen's Pride and Prejudice (BLOOM)
- 1340L Walden** (THOREAU)
- 1300L Arctic Dreams: Imagination and Desire in a Northern Landscape (LOPEZ)

1200L

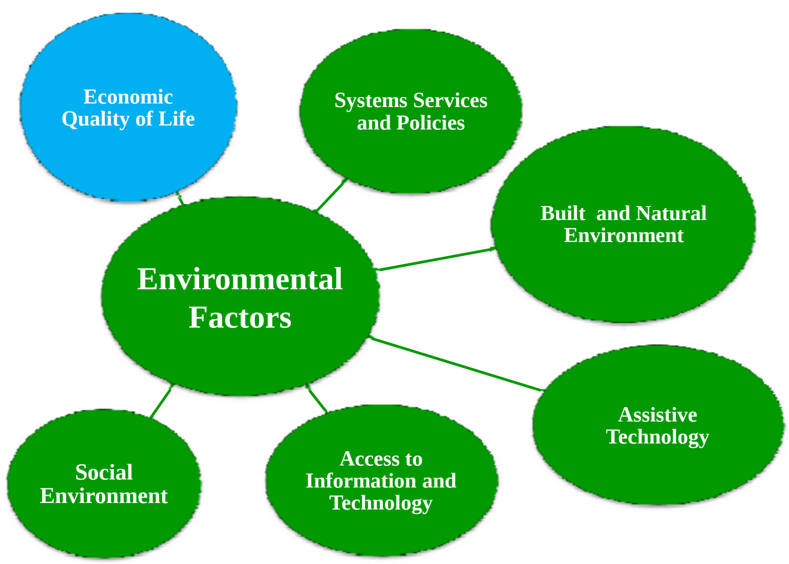
- 1280L The House of the Spirits (ALLENDE)
- 1270L Tarzan of the Apes (BURROUGHS)
- 1270L Chronicle of a Death Foretold (GARCÍA MÁRQUEZ)
- 1220L Annie John (KINCAID)

- 1290L A Brief History of Time (HAWKING)
- 1280L Black, Blue, and Gray: African Americans in the Civil War** (HASKINS)
- 1240L Blood Done Sign My Name (TYSON)

21 August 2014

Example
Measuring Environmental Factors for Persons with Stroke, Traumatic Brain Injury and Spinal Cord Injury

Conceptual Framework



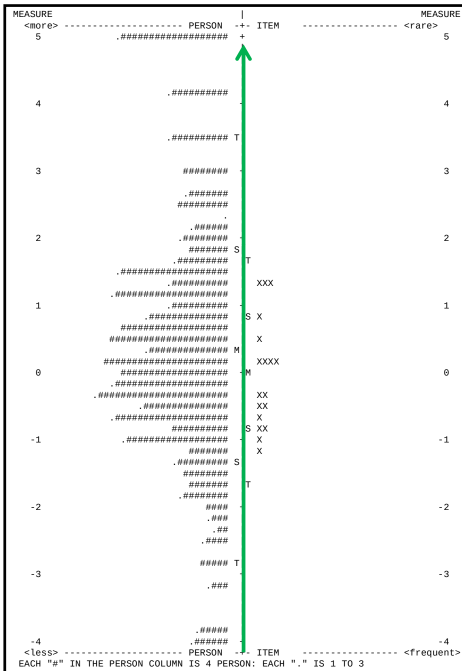
54

Comparison of ICF and Developed Conceptual Frameworks

ICF Taxonomy	Initial Framework	Binned Items	Winnowed Items	Revised Framework	Final Item Count
Products and Technology	Assistive technology	178	104	Assistive technology	15
Natural Environment and human made changes	Natural environment	79	37	Built and natural environment	36
	Built environment	605	51		
Support and Relationships	Social supports and attitudes	710	91	Social environment	82
Attitudes					
Services, Systems and Policies	Services, systems, policies	411	35	Services, systems, policies	75
	Transportation	136	28		
	Access to information and technology	112	37	Access to information and technology	25
	Economic-financial	42	35	Economic Quality of Life	32
Total		2,273	392		265

Economic quality of life: Rating scale analysis results

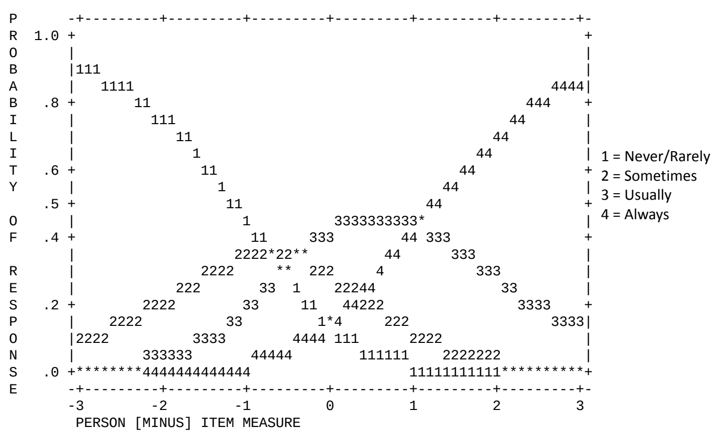
Analysis	Items Categories People	Person Item Reliability	Rating Scale	Misfitting Items (mn sq infit >1.4	Item-Measure r's <.4	PCA Residual Variance	Notes
1	24i 5c 1812p	.91 1.00	Monotonic	24 good credit 26 afford school 32 saved 6mo	None	5.6%	Combine response categories 1,2
2	24i 4c 1812p	.92 1.00	Monotonic	24 good credit 26 afford school 32 saved 6mo	None	5.2%	Delete 3 misfitting items and 3 over fitting items (6, 7, 10)
3	18i 4c 1812p	.90 1.00	Monotonic	None	None	6.0%	



Map of Economic QOL Items and Respondents

21 August 2014 57

Economic QOL Rating scale – measure map



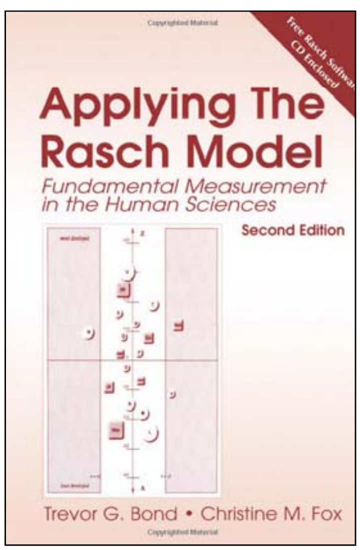
21 August 2014 58

Economic QOL item statistics in measure order

Measure	S.E.	Infit MnSq	Infit ZSTD	I-M Corr	ITEM
1.37	.04	.96	-1.2	.77	3 I have enough income to live the life that I want.
1.37	.04	1.23	5.9	.73	4 I can afford to live where I want to.
1.29	.03	.98	-.4	.75	11 I can afford to travel
.77	.03	1.35	9.2	.70	31 I have access to extra money in the case of an emergency
.55	.03	.82	-5.7	.77	2 I have enough income to meet my needs.
.24	.03	1.16	4.4	.71	18 I have opportunities to make enough money to live in my community
.20	.03	.82	-5.5	.75	9 I can afford to participate in community events.
.18	.03	1.15	4.3	.68	1 I can afford to eat out when I want.
.15	.03	1.22	6.1	.70	8 I can afford the personal care assistance that I need.
-.31	.03	1.18	4.9	.71	21 I can afford cable TV service.
-.35	.03	.80	-6.0	.75	5 I can afford to pay my bills.
-.45	.03	1.19	5.1	.69	20 I can afford internet service.
-.53	.03	.85	-4.5	.73	19 I can afford to buy healthy food.
-.73	.03	.99	-.2	.71	12 I have enough income to pay my rent or mortgage.
-.80	.04	.95	-1.5	.71	14 I can afford air conditioning for my home.
-.82	.04	.76	-7.3	.73	22 I can afford to pay my bills on time.
-1.04	.04	.91	-2.5	.71	13 I can afford heating my home.
-1.10	.04	1.15	3.7	.66	17 I can afford to feed myself and my family.

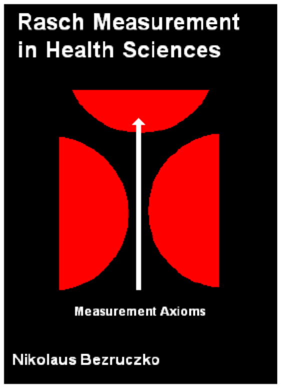
21 August 2014 59

Rasch Measurement Resources



21 August 2014 60

Rasch Measurement Resources

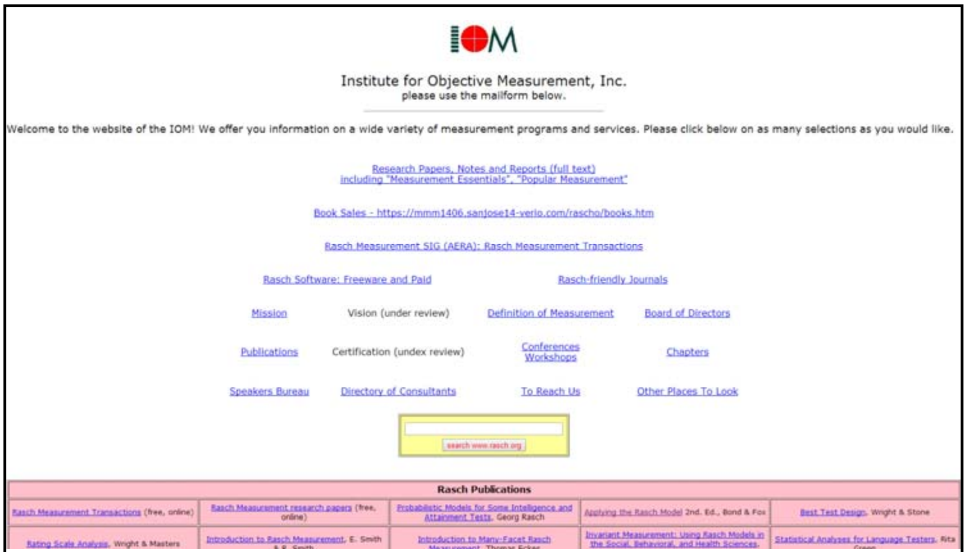


Rasch Measurement
in Health Sciences

Edited by
Nikolaus Bezruczko

JAM Press is pleased to announce the new book, *Rasch Measurement in Health Sciences*, is now available. The book is available in soft cover ($44, ISBN 0-9755351-3-7) and hard cover ($56, ISBN 0-9755351-2-9). Postage and handling are additional. Information on ordering the book is located on the reverse of this announcement. Information is also available at the *Journal of Applied Measurement* web site (www.jampress.org). Please go to the JAM Press Books page on the web site and scroll down to the new books section.

21 August 2014 61



www.rasch.org

21 August 2014 62

They may all fly (the construct)…
but, we won't know if birds of different feather can be calibrated together without verifying they "add up" to the same construct

Discussion

21 August 2014

64

Acknowledgments

RRTC on Improving Measurement
of Medical Rehabilitation Outcomes

Midwest Regional Spinal Cord Injury Care System

This presentation does not necessarily represent the policy of the Department of education, and you should not assume endorsement by the Federal Government.

21 August 2014

65