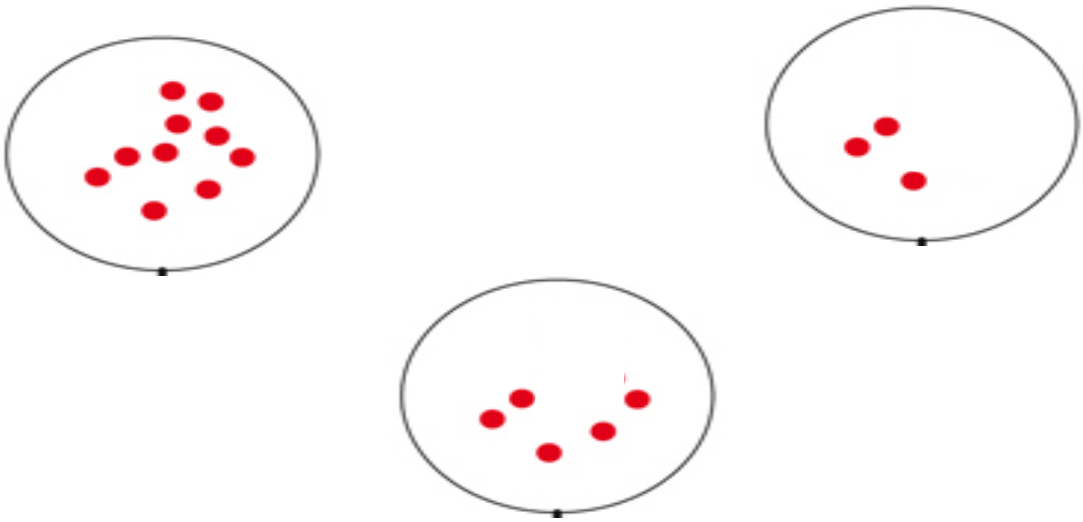
IRT or Rasch – what can and cannot be done?

L R Pendrill[1], and the EMPIR NeuroMet 15HLT04 consortium

[1]RI.SE Research Institutes in Sweden, Metrology, Eklandagatan 86, 41261 Göteborg (SE), phone:+46 767 88 54 44, mailto:lesliependrill@gmail.com

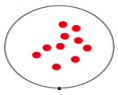
How many dots?



Man as Measurement Instrument: Counting

http://itre.cis.upenn.edu/~myl/CommuniqueMundurucuENG.pdf

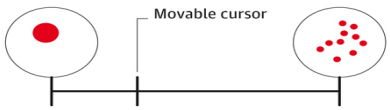
'*M* dots'

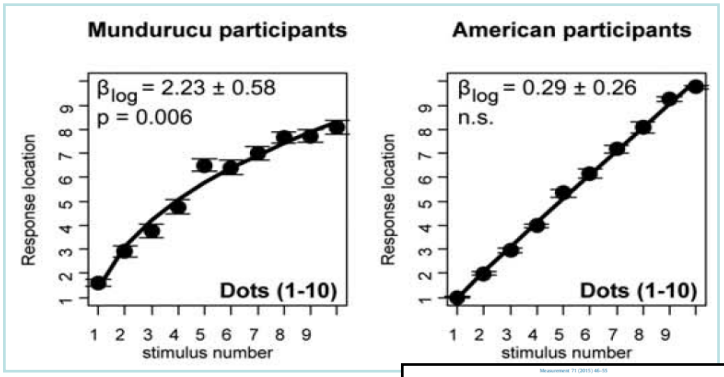


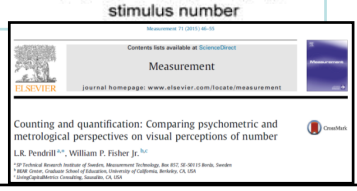
www.sciencemag.org SCIENCE VOL 320 30 MAY 2008

Log or Linear? Distinct Intuitions of the Number Scale in Western and Amazonian Indigene Cultures

Stanislas Dehaene,* Véronique Izard, Elizabeth Spelke, Pierre Pica

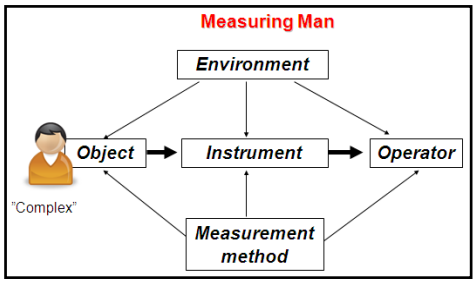


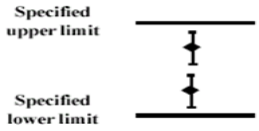


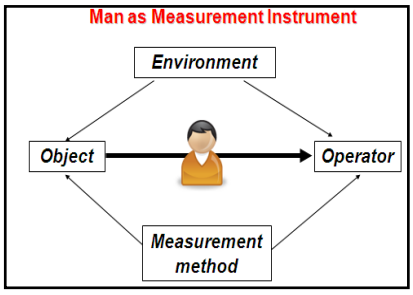


Lesliependrill@gmail.com 2019-06-12 PM-Health

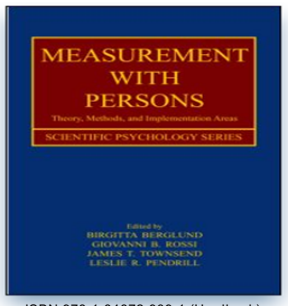
Man as Measurement Instrument







Measuring Man:
- Status, function of person
- Test against specifications



ISBN 978-1-84872-939-1 (Hardback)

Man as Measurement Instrument:
- Perception of product/service function, comfort etc
- Propose improvements in product

4

Lesliependrill@gmail.com 2019-06-12 PM-Health

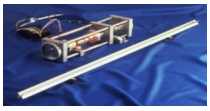
IRT or Rasch – what can and cannot be done?

- Man as Measurement Instrument
- Quality-assured measurement
- Logistic ruling & counted fractions
- Uncertainty & bias in measurement systems
- Classical test theory or Rasch?
- Aquiescence and Construct Alleys
- Rasch or 3PL IRT?

5

Quality-assured measurement
Metrology

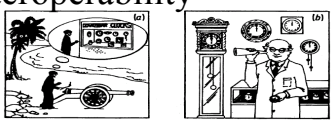
kg



m

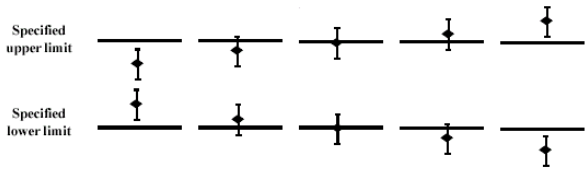
?

- Traceability ⇔ Comparability, Interoperability



"Zanzibar effect" [Harrison, MIT]

- Uncertainty ⇔ Risks



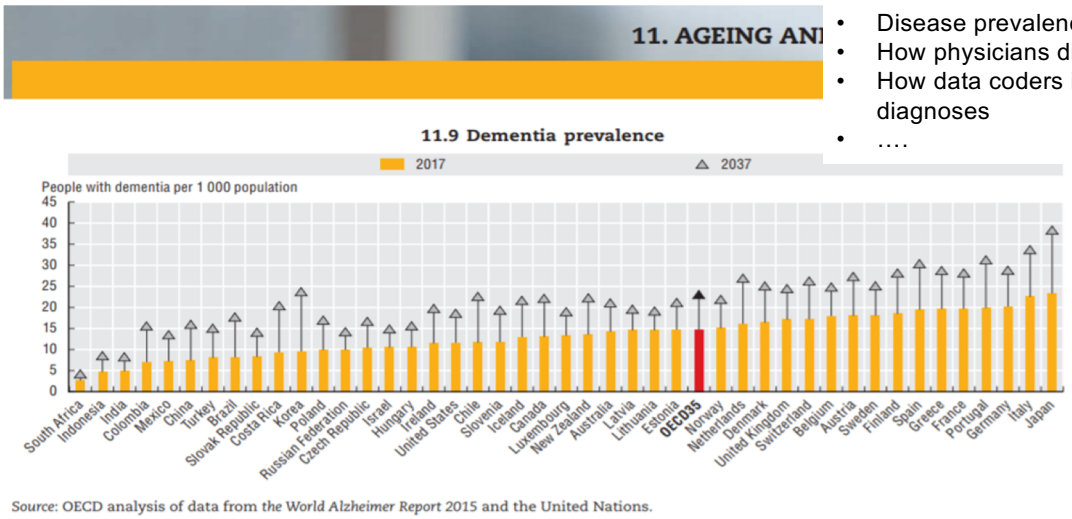
1,234 m ± 0,016 m

6

An (un)reasonable difference?

Potential causes of variation:

- Disease prevalence
- How physicians diagnose
- How data coders interpret diagnoses
- ….



Source: OECD analysis of data from *the World Alzheimer Report 2015* and the United Nations.

7

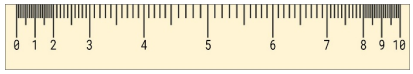
Quality-assured categorical measurement

”Satisfaction”

8

http://www.mynewsdesk.com/se/region_skane/news/patienter-vaerderar-vaarden-med-en-knapptryckning-143455

https://www.alz.org/alzheimers-dementia/diagnosis/medical_tests

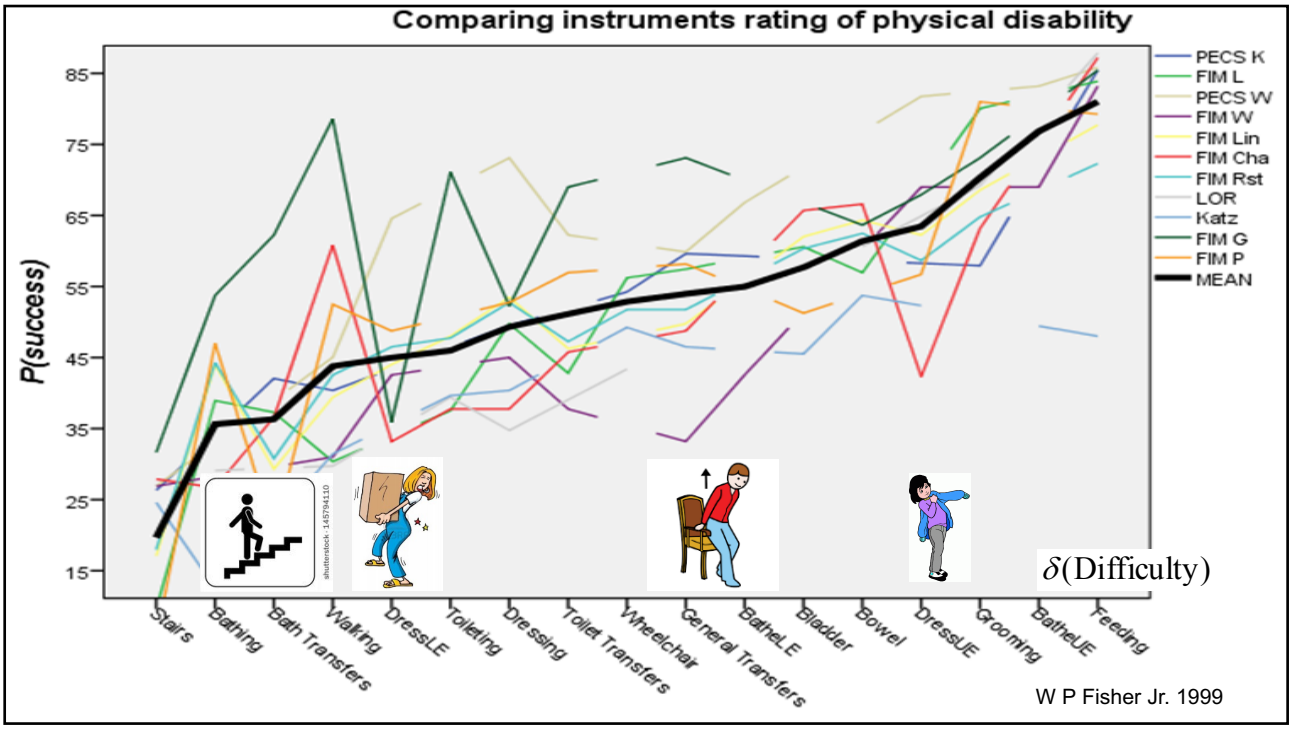


δ: Difficulty, quality

θ: Ability, sensitivity, leniency

9

Lesliependrill@gmail.com 2019-06-12 PM-Health



W P Fisher Jr. 1999

Tasks

$\delta(Difficulty)$

Physical disability

Difficulty

Metrological references

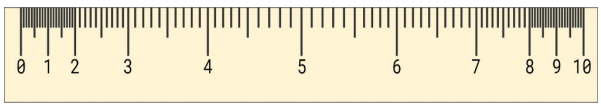
Mass

11

Different scales of measurement

$x \geqq y$ iff $\phi(x) \geqq \phi(y)$ (Strictly) monotone increasing transformation	Ordinal	Preference? Hardness Air quality Grades of leather, lumber, wool, etc. Intelligence tests, raw scores

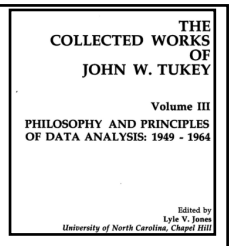
$$1+2+3 \approx 6$$



'Counted fractions'

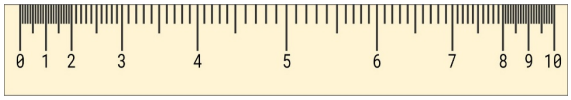
12

Counted fractions



Relative-number problems:

- Counting (sheep & goats)
- How many affected at this dose
- How many of the pebbles are quartz…



$$X_j\% = \frac{X_j}{\sum_{k=1}^{K} X_k}\%$$

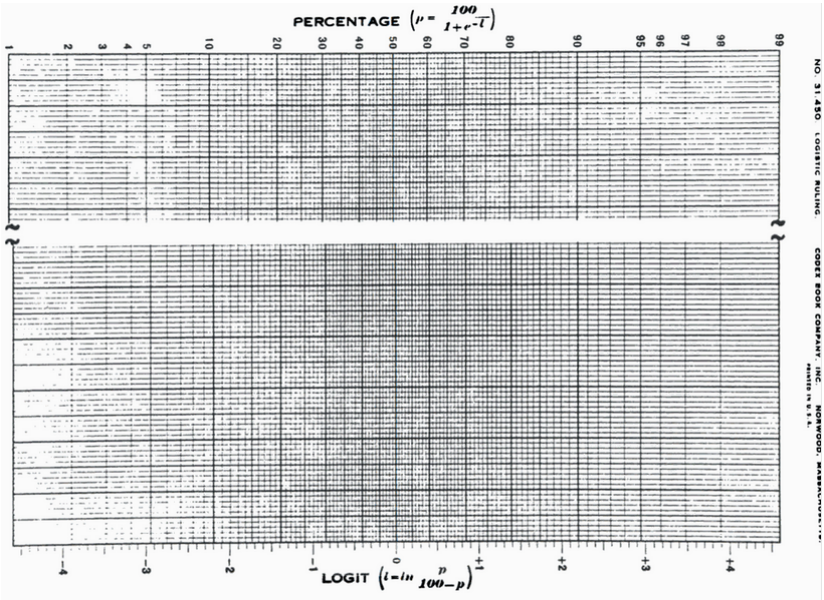
X_j X_j

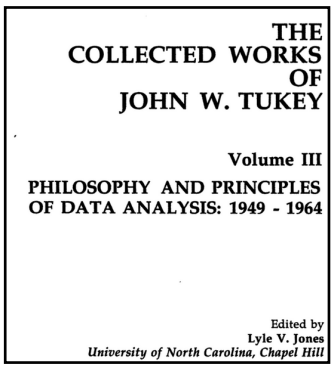
$$1+2+3 \approx 6$$

"Beware of attempts to interpret correlations between ratios whose numerators and denominators contain common parts"
[Pearson 1897]

Tukey [Chapter 8, Data analysis and behavioural science; quoted in "The collected works of John A Tukey, Volume III, Philosophy and principles of data analysis: 1949 – 1964", ed. L V Jones, Univ. North Carolina, Chapel Hill

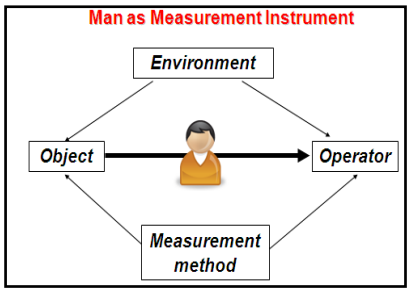
Logistic ruling



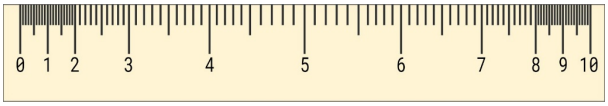


$$\log\left(\frac{P_{success}}{1-P_{success}}\right) = z$$

'Counted fractions'



$$Y = P_{success}, response$$



Counted fractions (Rasch)

$$Y\% = \frac{Y_j}{\sum_k Y_k};\ \sum_k Y_k = 100\%$$

$$P_{success} = \frac{e^{\theta-\delta}}{1+e^{\theta-\delta}}$$

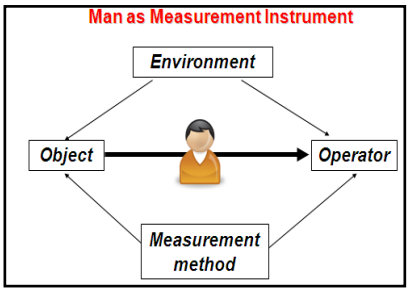
θ – instrument (person) ability

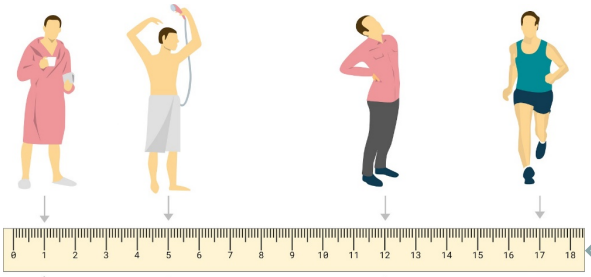
δ – object (task) difficulty

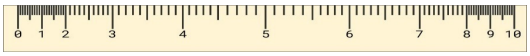
15

θ = *person ability*

Measuring People







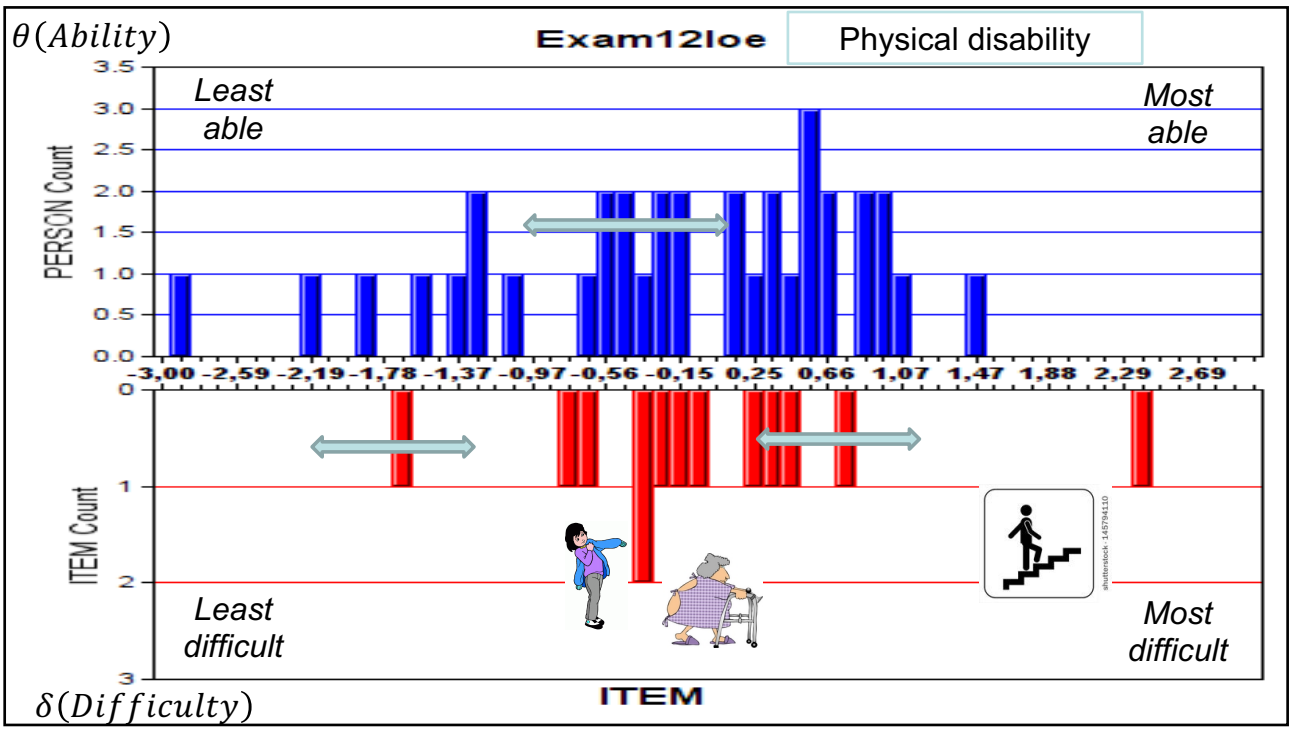
$$log\left(\frac{P_{success}}{1-P_{success}}\right) = \theta - \delta$$

Very difficult	Difficult	Easy	Very easy
☐	☐	☐	☐

δ = *object difficulty*

Rasch (1961)

16

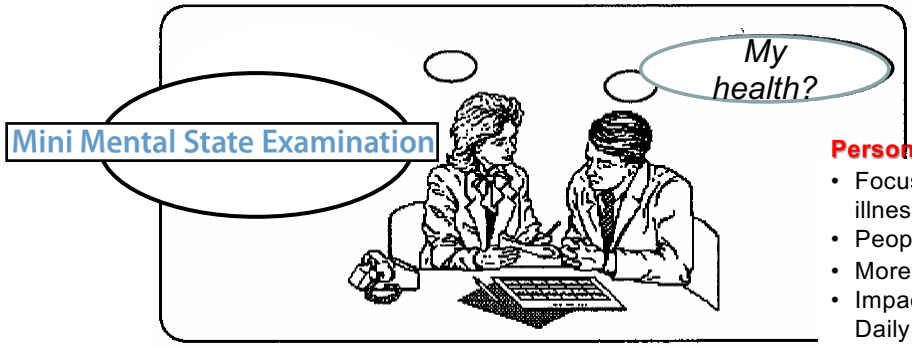

θ(Ability)
Exam12loe
Physical disability
Least able
Most able
PERSON Count
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
-3,00 -2,59 -2,19 -1,78 -1,37 -0,97 -0,56 -0,15 0,25 0,66 1,07 1,47 1,88 2,29 2,69
ITEM Count
0
1
2
3
Least difficult
Most difficult
ITEM
δ(Difficulty)


IOP Publishing
Measurement Science and Technology
Meas. Sci. Technol. 29 (2018) 034003 (13pp)
https://doi.org/10.1088/1361-6501/aa9cd2
Assuring measurement quality in person-centred healthcare
L R Pendrill
[1] National Metrology Institute, Research Institutes of Sweden, Eklandagatan 86, 412 61 Göteborg, Sweden
E-mail: leslie.pendrill@ri.se
Received 30 June 2017, revised 26 September 2017
Accepted for publication 23 November 2017
Published 8 February 2018
CrossMark
SPECIAL FEATURE
Man as a Measurement Instrument
Leslie Pendrill
24 | NCSLI Measure J. Meas. Sci.
www.ncsli.org
IMEKO2016 TC1-TC7-TC13
Journal of Physics: Conference Series 772 (2016) 012025
IOP Publishing
doi:10.1088/1742-6596/772/1/012025
On Trial: the Compatibility of Measurement in the Physical and Social Sciences
S J Cano[1], T Vosk[2], L R Pendrill[3], A J Stenner[4]
[1]Modus Outcomes, Spirella Building, Letchworth Garden City, SG6 4ET, UK, stefan.cano@modusoutcomes.com; [2]8105 NE 140th Pl, Kirkland, WA 98034, tvosk@comcast.net; [3]SP Technical Research Institute of Sweden, Metrology, Box 857, SE-50115 Borås, Sweden, leslie.pendrill@sp.se; [4]MetaMetrics, Inc., 1000 Park Forty Plaza Drive, Suite 120, Durham, NC 27713, USA & The University of North Carolina, Chapel Hill, North Carolinam USA, jstenner@Lexile.com.
Measurement 71 (2015) 46–55
Contents lists available at ScienceDirect
Measurement
journal homepage: www.elsevier.com/locate/measurement
ELSEVIER
Counting and quantification: Comparing psychometric and metrological perspectives on visual perceptions of number
CrossMark
L.R. Pendrill [a,*], William P. Fisher Jr. [b,c]
[a] SP Technical Research Institute of Sweden, Measurement Technology, Box 857, SE-501 15 Borås, Sweden
[b] BEAR Center, Graduate School of Education, University of California, Berkeley, CA, USA
[c] LivingCapitalMetrics Consulting, Sausalito, CA, USA
Patient-centred outcome metrology for healthcare decision-making
S J Cano[1,5], L R Pendrill[2], S P Barbic[3] and W P Fisher Jr[4]
[1] Modus Outcomes, Spirella Building, Letchworth Garden City, SG6 4ET, UK
[2] RLSE, Eklandagatan 86, S-412 61, Göteborg, Sweden
[3] Faculty of Medicine, Department of Occupational Science and Occupational Therapy, the University of British Columbia, Vancouver, BC V6T 2A1, Canada
[4] BEAR Center, Graduate School of Education, University of California, Berkeley, CA 92720 USA
18
Lesliependrill@gmail.com 2019-06-12 PM-Health

19

Quality-assured measurement

Object: Health



Person-centred care (PCC)

- Focus on health (not illness)
- People partners in care
- More symptoms
- Impact on Activities of Daily Living
- Subjective & perceptive
- …

Cognitive ability?

0,8 units ± 0,2 units

20

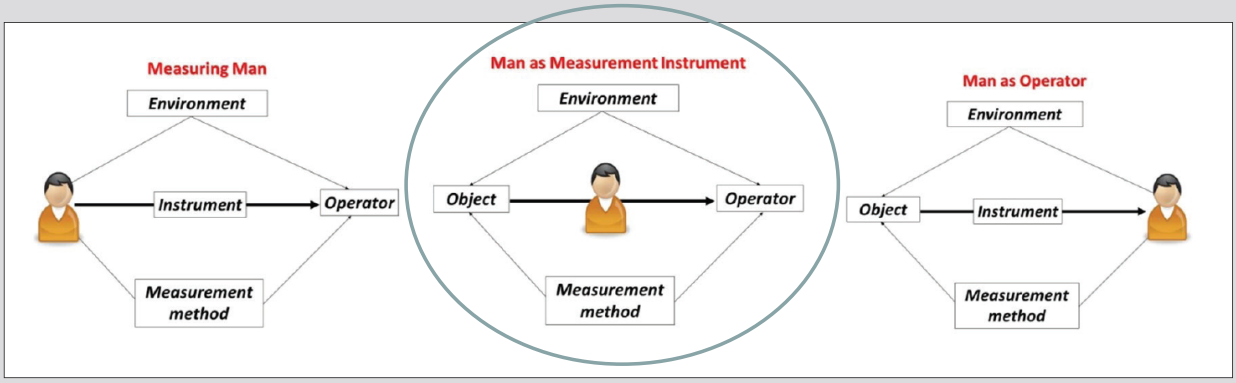
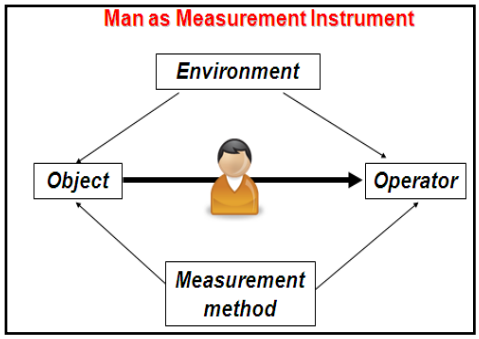


Figure 1. Different human interventions in measurement systems.

L R Pendrill 2014, "Man as a Measurement Instrument".
NCSLI Measure J. Meas. Sci., **9**: 24 – 35.

21

Possible way forward equally applicable to both physical and social measurement




IOP Publishing
Measurement Science and Technology
Meas. Sci. Technol. 00 (2016) 000000 (17pp)
UNCORRECTED PROOF
Metrology of human-based and other qualitative measurements
Leslie Pendrill and Niclas Petersson
SP Technical Research Institute of Sweden, Measurement Technology, Box 857, S-501 15 Borås, Sweden
E-mail: xxxx
Received 20 November 2015, revised 7 January 2016
Accepted for publication 8 January 2016
Published

Physics & engineering ⇄ Social sciences

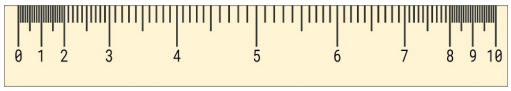
Object value	Confidence interval	Decision risks	"Vagueness"
1.234 m	1.234 (± **0.012**) m	$\mathbf{P} = \begin{pmatrix} 1-\alpha & \alpha \\ \beta & 1-\beta \end{pmatrix}$	?

22

‘Instruments’

Measuring instrument: “device used for making measurements, alone or in conjunction with one or more supplementary devices” (§3.1 VIM)

Social sciences
- Rating scales, questionnaires, ability tests,…



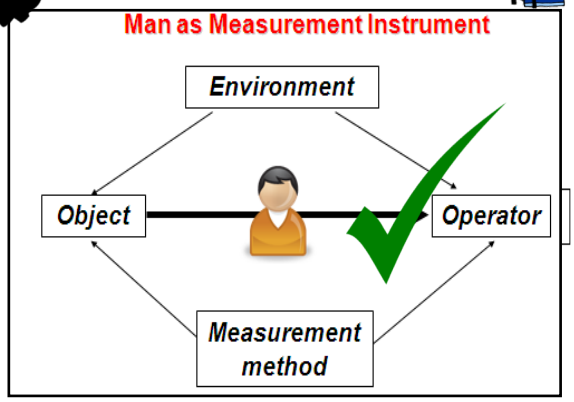
Physics and engineering
- Rulers,voltmeters, LIGO,…

LIGO Laser Interferometer Gravitational-Wave Observatory
Supported by the National Science Foundation
Operated by Caltech and MIT

23

‘Instruments’

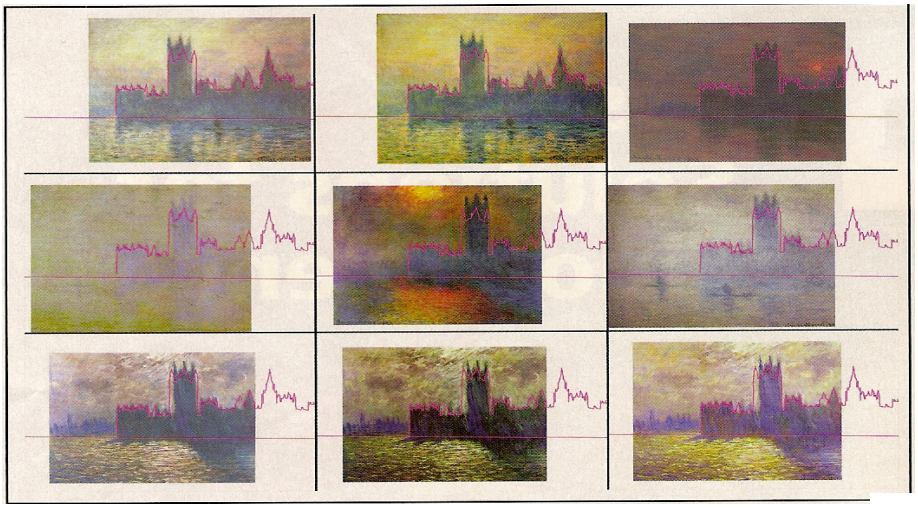
Social sciences
- Rating scales, questi ires, ability tests,…



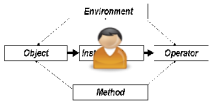
‘Instruments’

24

Man as Measurement Instrument



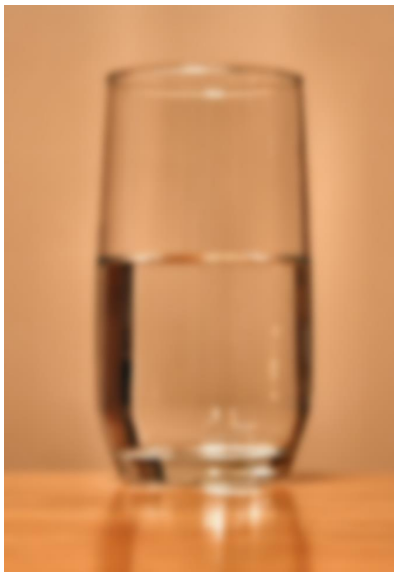
Monet



25

Lesliependrill@gmail.com 2019-06-12 PM-Health

http://www.cartage.org.lb/en/themes/arts/drawings/Silhouettes/CreatingMe/monet1.jpg



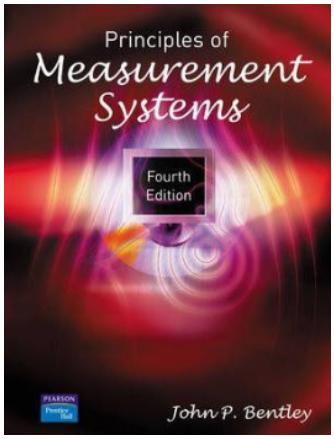
Uncertainty

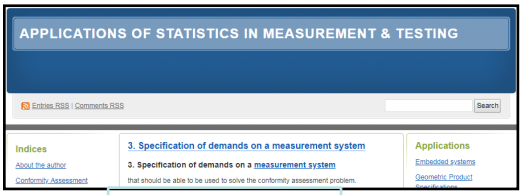
Bias

26

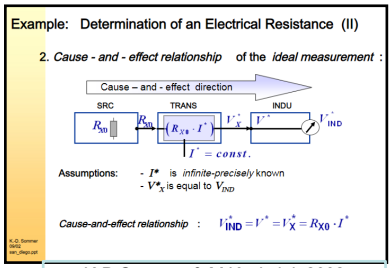
Lesliependrill@gmail.com 2019-06-12 PM-Health

Measurement systems

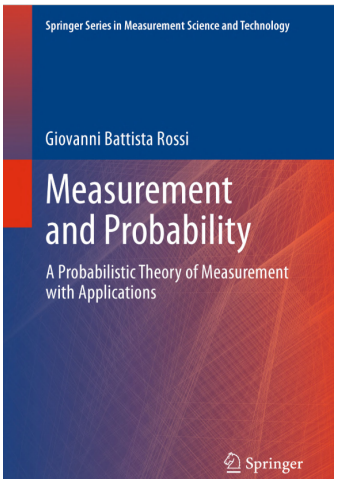




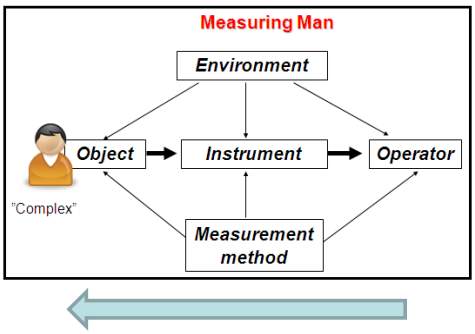
Metrology.wordpress.com



K-D Sommer & M Kochsiek 2002



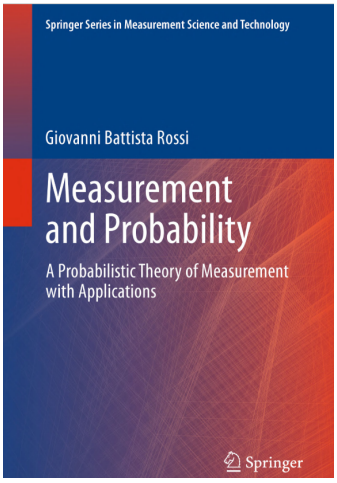
27



"Measurement value":

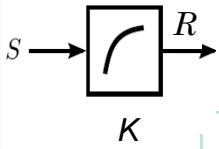
≠ indication of **response** of measurement system

= **restituted** value.

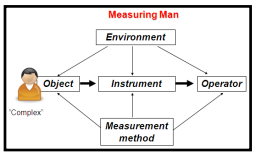


z_R – restituted value (measurement result)

28



Measurement systems



Balance as Measurement Instrument - *Sensitivity (**K**), bias (**b**)*

*Stimulus (**S**)*: Mass of weight

Calibration
$R_{cal} = K_{cal} \cdot S_{cal} + b$

$R = K \cdot S$ + "additional terms"

*Response (**R**)*:

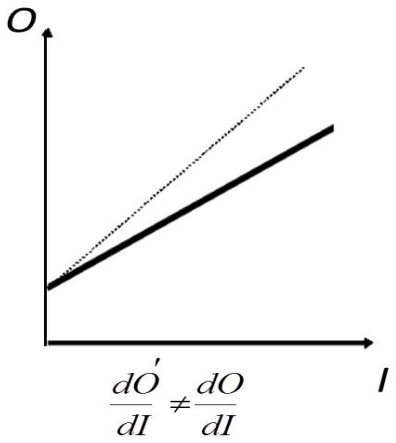
Mass of weight x
Balance sensitivity

Measurand 'restitution', $S = \frac{R-b}{K_{cal}}$

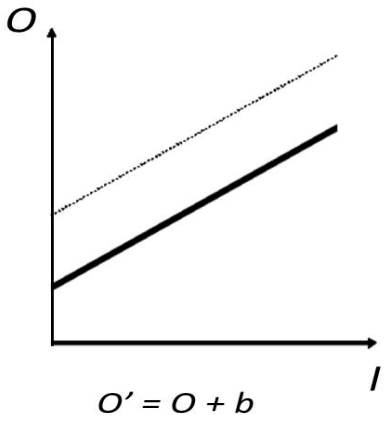
Measurement systems

INPUT		OUTPUT
I_{MIN} to I_{MAX}	*RANGE*	O_{MIN} to O_{MAX}
I_{MIN} - I_{MAX}	*SWING*	O_{MIN} - O_{MAX}
I	*NON-LINEARITY*	$O = f(I)$
I	*SENSITIVITY*	$K = \frac{dO}{dI}$
dI	*RESOLUTION*	$dI_{res} = dI_{max}$ when $dO = 0$
I	*WEAR*	$O(t_2) \neq O(t_1)$
I	*ERROR LIMITS*	$O - h < O_0 < O + h$
	ENVIRONMENTAL EFFECTS	$O' = f'(I)$
I	(a) Modifying	$K' = \frac{dO'}{dI} \neq \frac{dO}{dI}$
	(b) Interfering	$O' = O + b$
I_{up}, I_{down}	*HYSTERESIS*	$O(I_{upp}) - O(I_{ner})$

Measurement systems



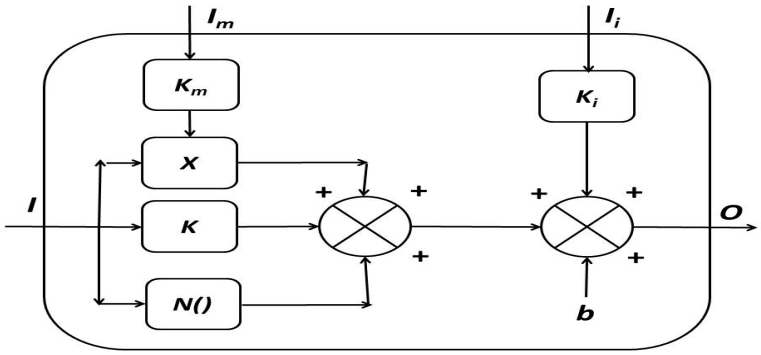
(a) Modifying



(b) Interfering

31

Measurement systems



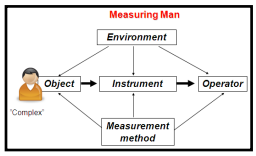
$$O = K \cdot I + N(I) + K_M \cdot I_M \cdot I + K_i \cdot I_i + b$$

Sensitivity = K; Non-linearity = $N(I)$; bias = b

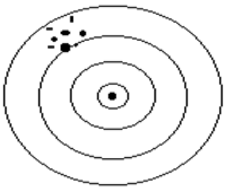
Modifying disturbance = I_M, with sensitivity = K_M; Interfering disturbance = I_i, with sensitivity = K_i.

32

Measurement systems



I → Sensing → Signal conditionning → Signal processing → Data presentation → *O*



Low trueness
High precision

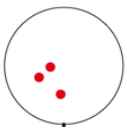
Accuracy (trueness)
= measured value – true value
= system output – system input, $O_i - O_{i-1} = O_i - I_i$

Accuracy (precision): $\sigma^2_{I_{i-1}} = \sigma^2_{O_i} = \left(\frac{\partial O_i}{\partial I_i}\right)^2 \cdot \sigma^2_{I_i} + \left(\frac{\partial O_i}{\partial I_{M_i}}\right)^2 \cdot \sigma^2_{I_{M_i}} + \cdots$

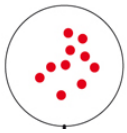
$Count\ 1: Bias = Classification - Actual = +0{,}5(6) \cdot [I]$

Expected response:

$$\mathbb{E}(y|z_M) = \sum_{k=1}^{K} \frac{P(y_k|z_M) \cdot y_k}{K}$$

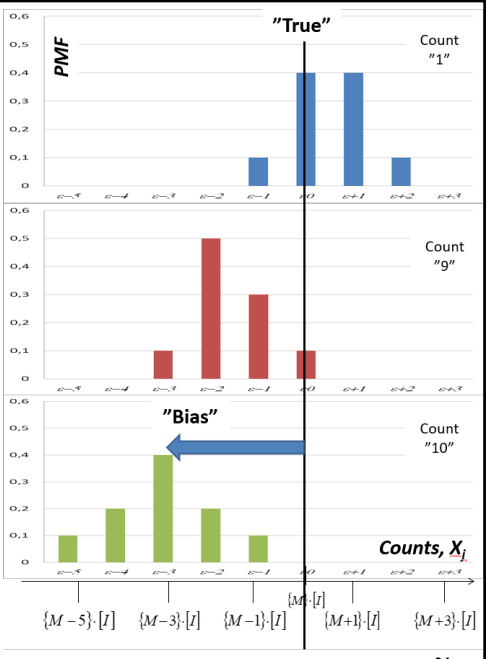


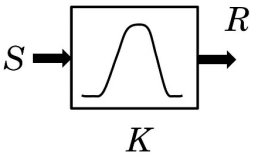
$Count\ 9: Bias = Classification - Actual = -1{,}7(6) \cdot [I]$



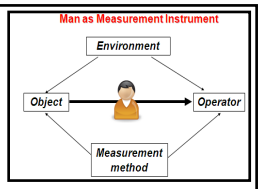
$Count\ 10: Bias = Classification - Actual = 3{,}1(8) \cdot [I]$

http://itre.cis.upenn.edu/~myl/CommuniqueMundurucuENG.pdf





Measurement systems



Man as Measurement Instrument - *Sensitivity (**K**)*

$\delta(Difficulty)$

*Stimulus (**S**)*: Task difficulty

$\theta(\text{Ability})$

***R** = **K**·**S** + "additional terms"*

$$P_{success} = \frac{e^{\theta-\delta}}{1+e^{\theta-\delta}}$$

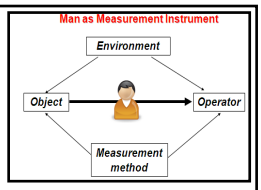
*Response (**R**)*:

Task difficulty x
'Instrument' sensitivity

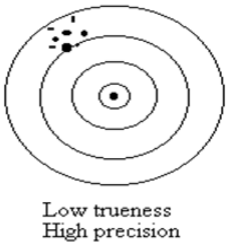
Measurand 'restitution', S = K^{-1}·R?

35

Measurement systems



I → Sensing → Signal conditionning → Signal processing → Data presentation → ***O***



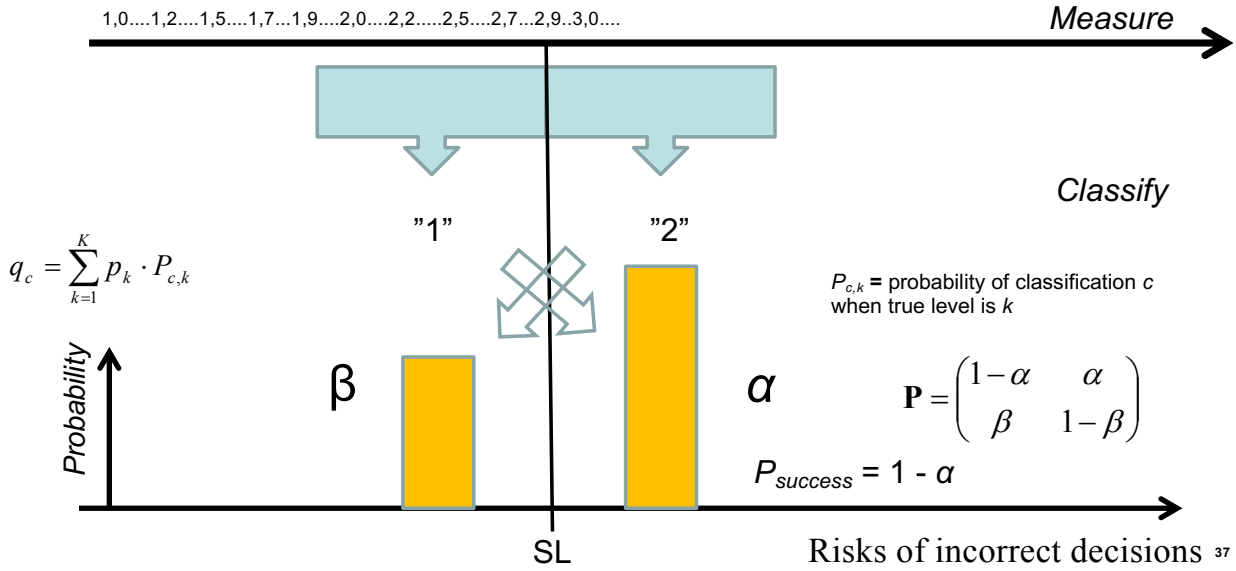
Accuracy (decision-making) =

response categorisation – input (true) categorisation

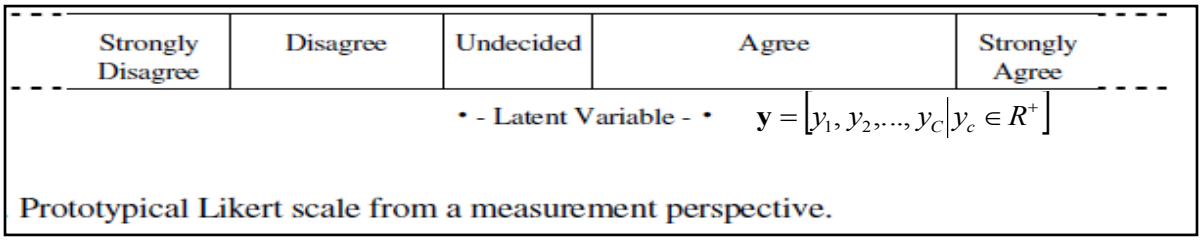
36

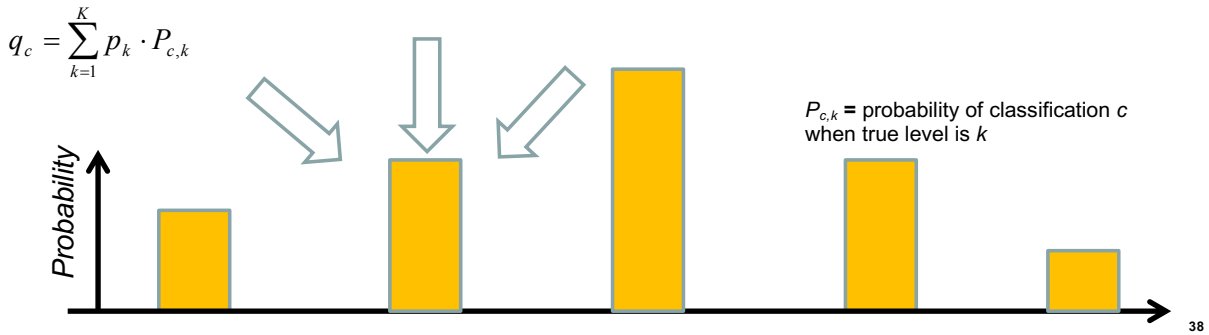
Binomial classification: *2* categories

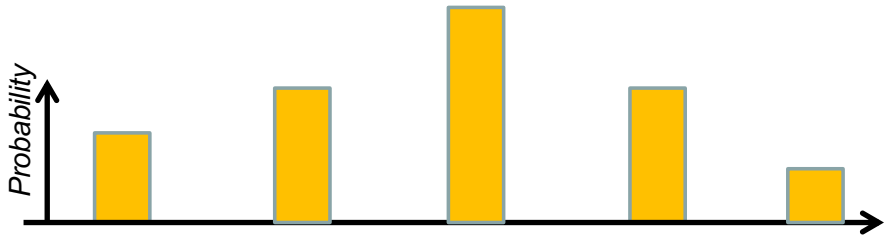
p_k = *a priori* probability that true level is *k*



p_k = *a priori* probability that true level is *k*







For categorical responses,
measurement system 'accuracy' = decision-making ability:

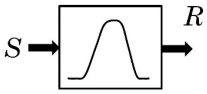
Accuracy (decision-making) = response categorisation – input (true) categorisation

where $P_{success}$ is metric of measurement system performance

in terms of probability of making 'correct' decision.

Measurement system analysis

Quantitative



Qualitative

Measurand
(stimulus, ***S***
e.g. task difficulty, δ, u_δ)

Response
indication, ***R***
e.g. $P_{success}$

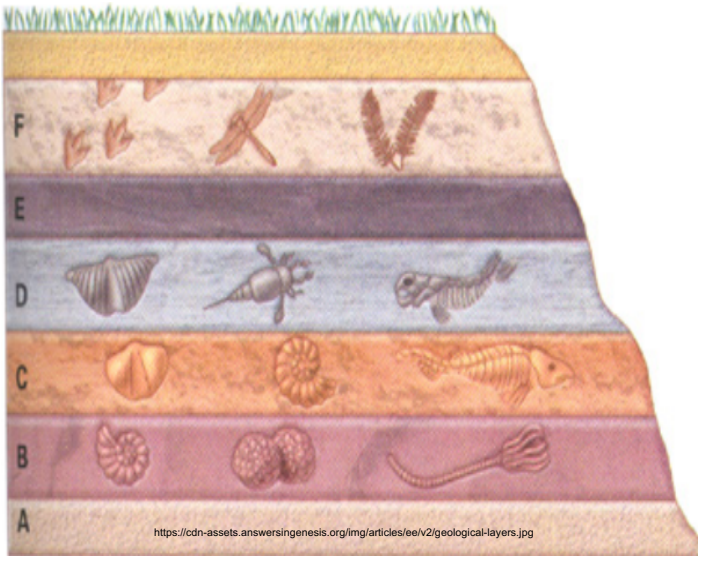
Restitution
R* => *S
e.g. $P_{success}$ => δ, u_δ

$$\log\left(\frac{P_{success}}{1-P_{success}}\right) = \theta - \delta$$

L R Pendrill 2018 *Meas. Sci. Technol.* https://doi.org/10.1088/1361-6501/aa9cd2

40

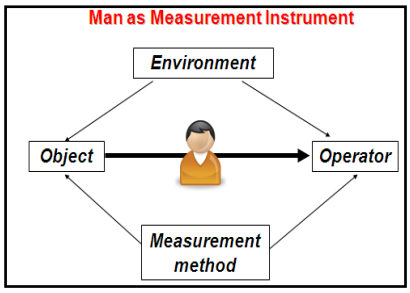
41

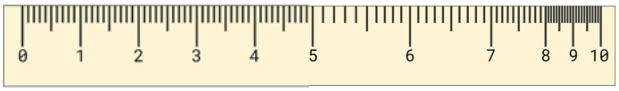


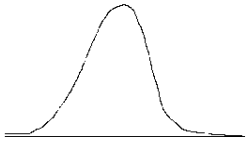
Measured
- depth (/m)

Required
- epoch
(/Myear)

42







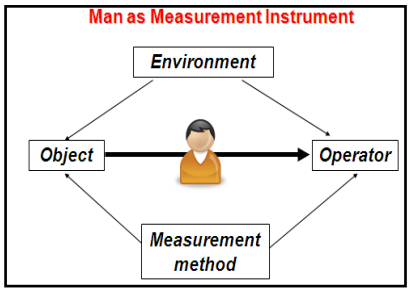
Acquiescence

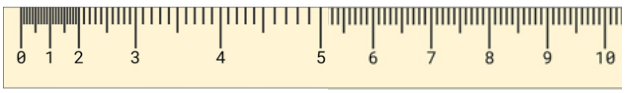
”Agreement regardless of item content

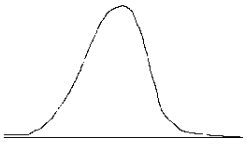
$$P_{success} = \frac{e^{\theta-\delta}}{1+e^{\theta-\delta}}$$

43

Lesliependrill@gmail.com 2019-06-12 PM-Health







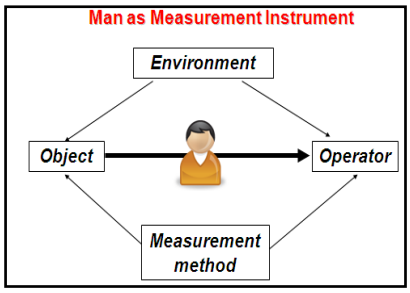
Disacquiescence

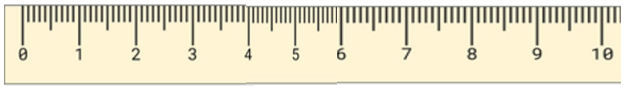
”Disagreement regardless of item content”

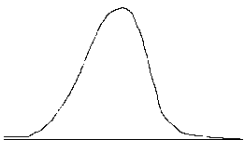
$$P_{success} = \frac{e^{\theta-\delta}}{1+e^{\theta-\delta}}$$

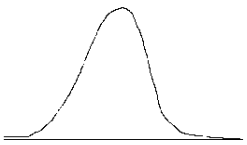
44

Lesliependrill@gmail.com 2019-06-12 PM-Health







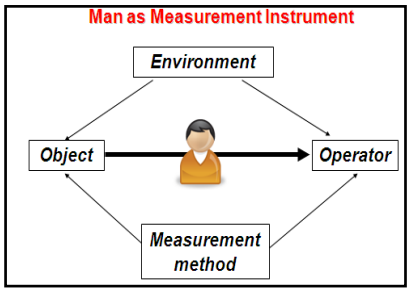


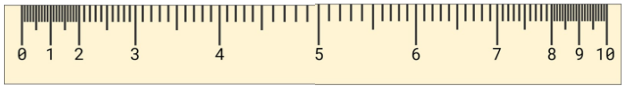
Extreme response bias

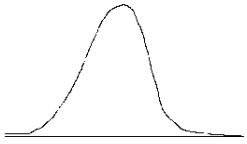
"Use scale endpoints regardless of item content"

$$P_{success} = \frac{e^{\theta-\delta}}{1+e^{\theta-\delta}}$$

45





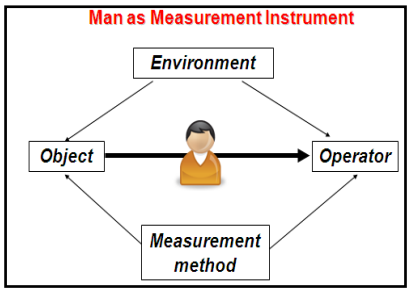


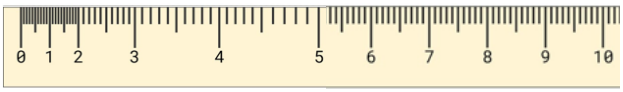
Middle response bias

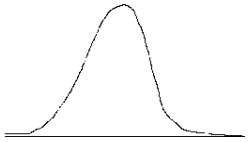
"Use scale midpoint regardless of item content"

$$P_{success} = \frac{e^{\theta-\delta}}{1+e^{\theta-\delta}}$$

46







Social desirability bias

”Present oneself in a positive way, regardless of item content”

$$P_{success} = \frac{e^{\theta-\delta}}{1+e^{\theta-\delta}}$$

47

The Developmental Psychometrics of Big Five Self-Reports: Acquiescence, Factor Structure, Coherence, and Differentiation From Ages 10 to 20

Christopher J. Soto and Oliver P. John, University of California, Berkeley
Samuel D. Gosling, University of Texas at Austin
Jeff Potter, Cambridge, Massachusetts

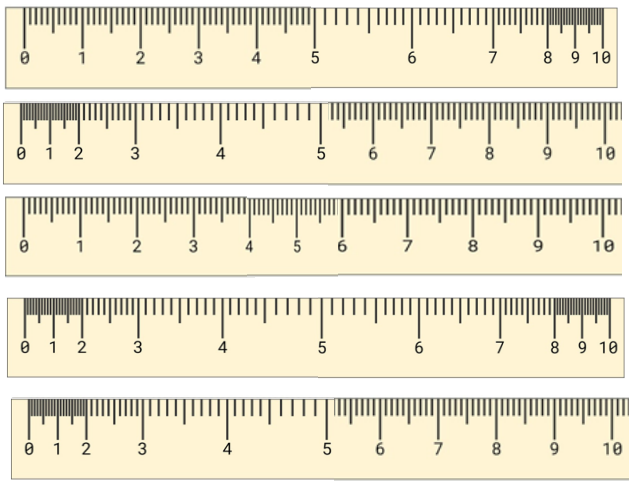
Acquiescence

Disacquiescence

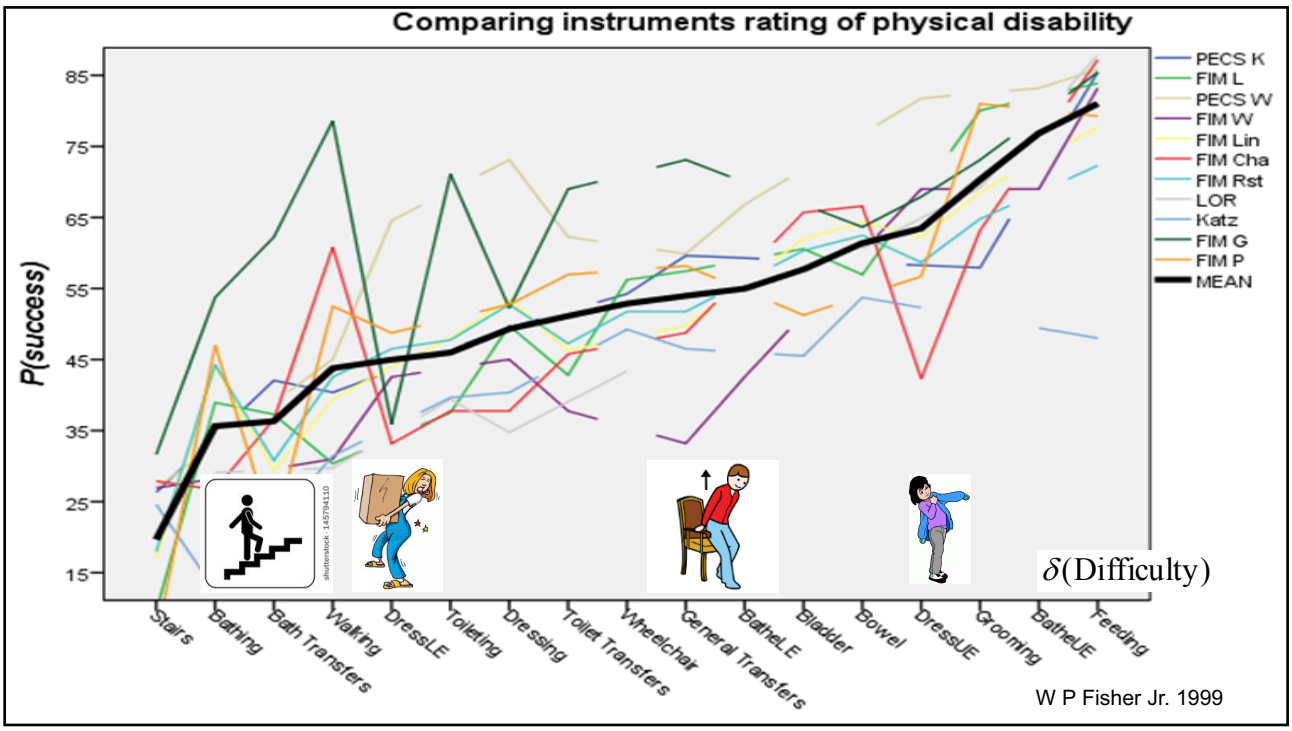
Extreme response bias

Middle response bias

Social desirability bias



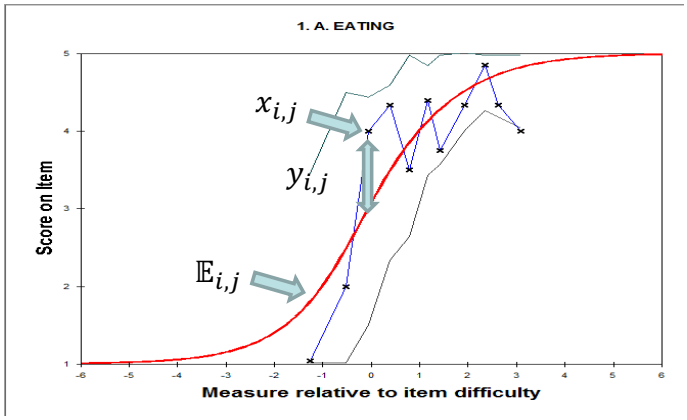
48

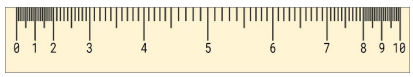


Score residual, *y* $y_{i,j} = x_{i,j} - \mathbb{E}_{i,j}$

Expected mean, *E*, of response *x* $\mathbb{E}_{i,j} = \sum_{k=0}^{C_j} k \cdot q_{i,j,k}$

$$q_{i,j,k} = \frac{e^{\sum_{c=0}^{k}(\theta_i - \delta_{j,c})}}{\sum_{k=0}^{C_j} e^{\sum_{c=0}^{k}(\theta_i - \delta_{j,c})}}$$





Graphs: Test Characteristic Curve

Exp+Empirical ICC

50

Lesliependrill@gmail.com 2019-06-12 PM-Health

3PL IRT or Rasch?

Workshop

51

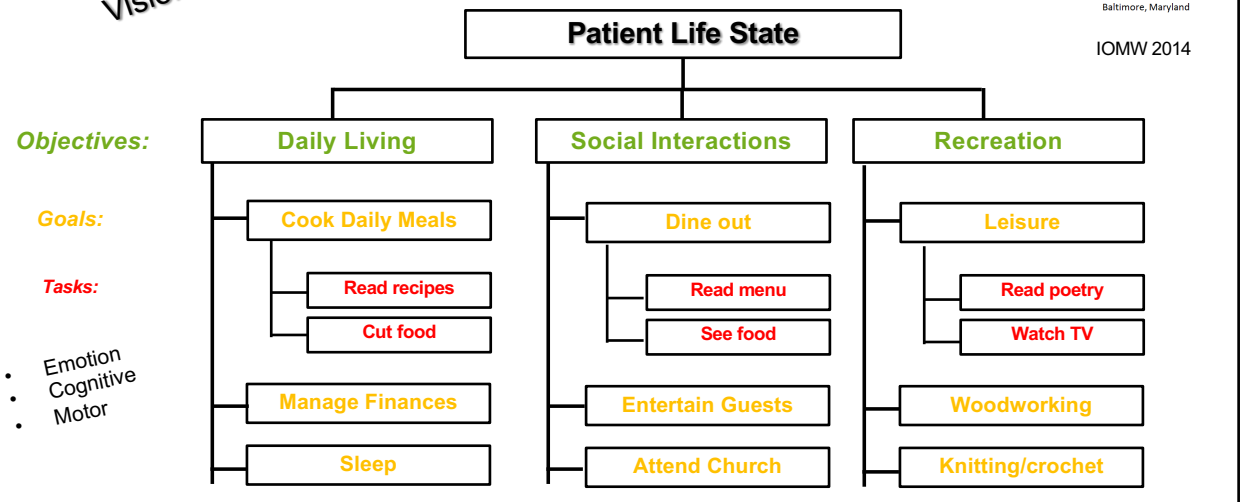
Activities – patient preferences

Are Subscales Compatible with Univariate Measures?

Robert W. Massof, Ph.D.
Johns Hopkins University
Baltimore, Maryland

IOMW 2014

Vision



52

Mean square fit statistic transformed to z-score (standard normal deviate)

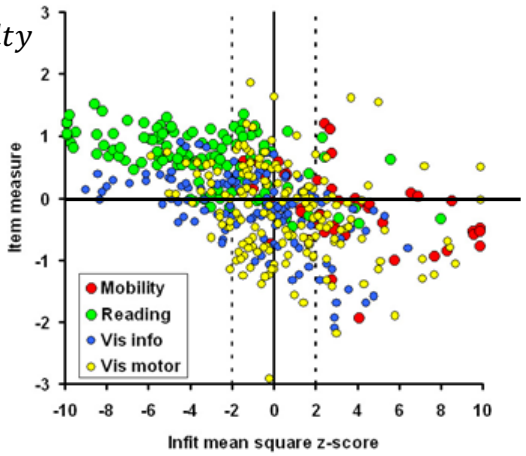
Are Subscales Compatible with Univariate Measures?

Robert W. Massof, Ph.D.
Johns Hopkins University
Baltimore, Maryland

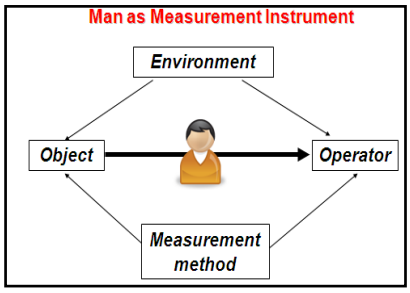
IOMW 2014

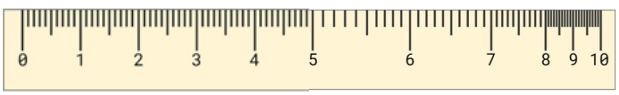
$\delta - difficulty$

- Emotion
- Cognitive
- Motor

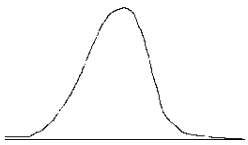


53





$\Delta\delta = \tau \cdot \delta; \tau < 0$



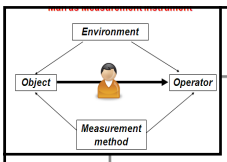
Acquiescence

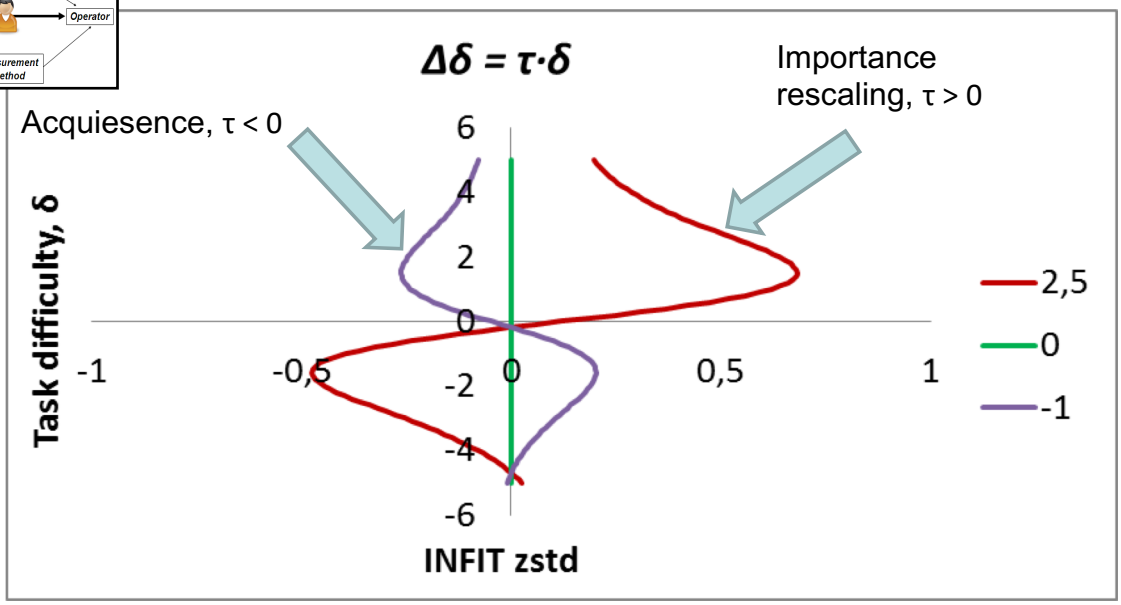
"Agreement regardless of item content

$$P_{success} = \frac{e^{\theta-\delta}}{1+e^{\theta-\delta}}$$

54

54

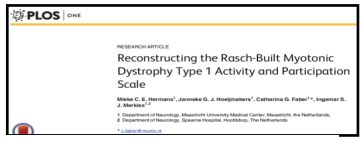




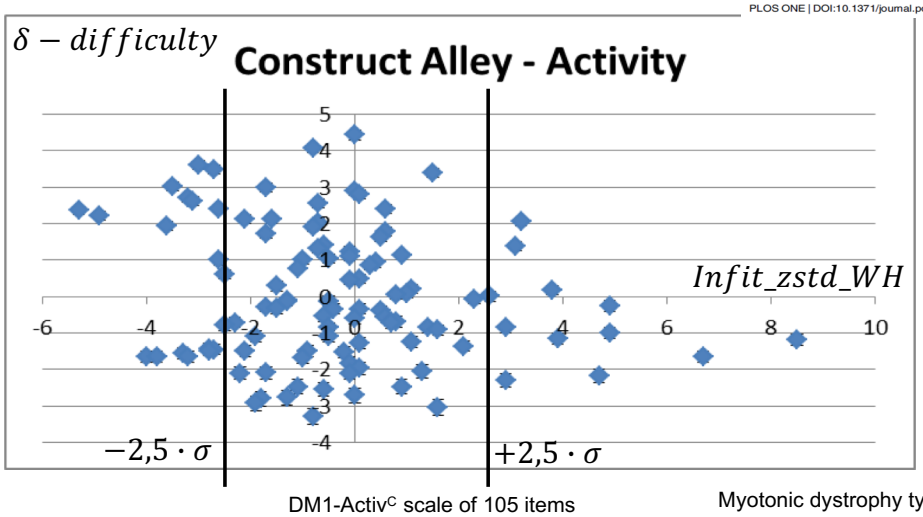
55

Fitting data to the Rasch model

The individual item fit statistics of 32 items demonstrated model misfit (having significant chi-square probability and/or fit residuals exceeding ±2.5) and were removed one by one (73 items remaining).


PLOS ONE | DOI:10.1371/journal.pone.0139944 October 20, 2015 1 - 9

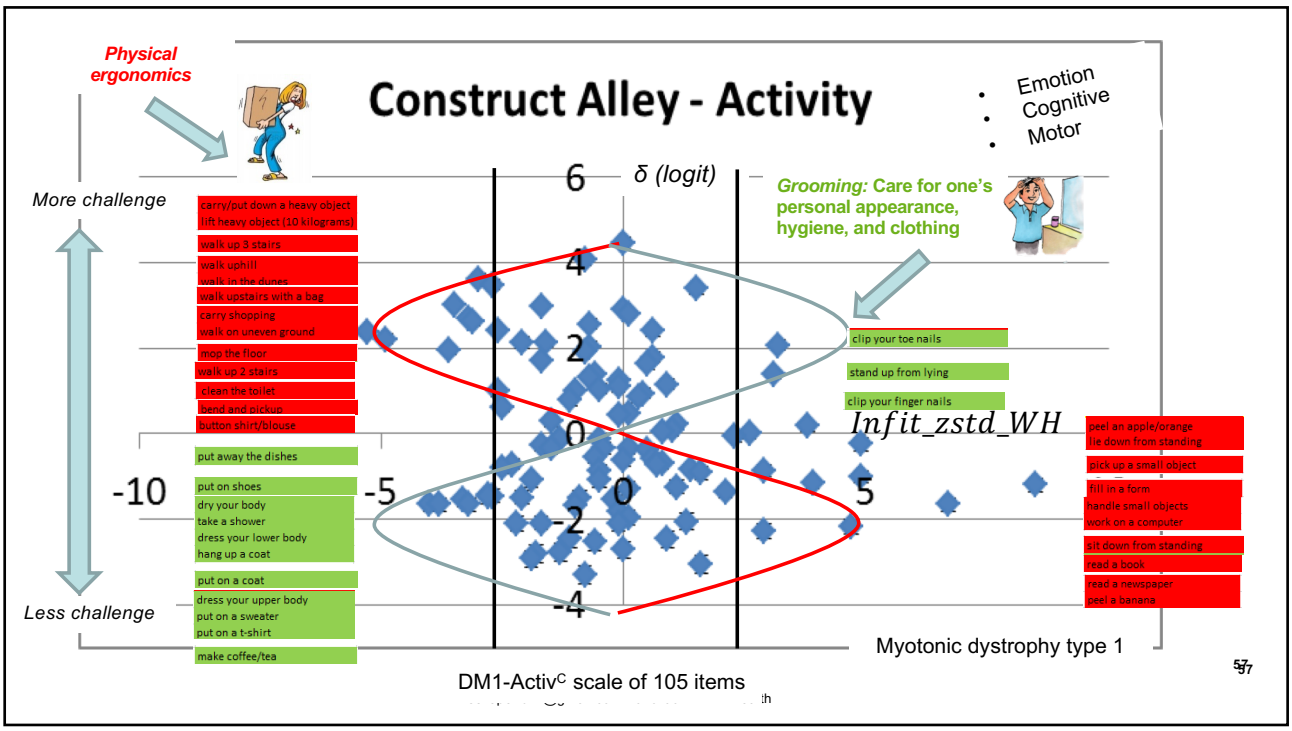
Reconstructing the Rasch-Built Myotonic Dystrophy Type 1 Activity and Participation Scale

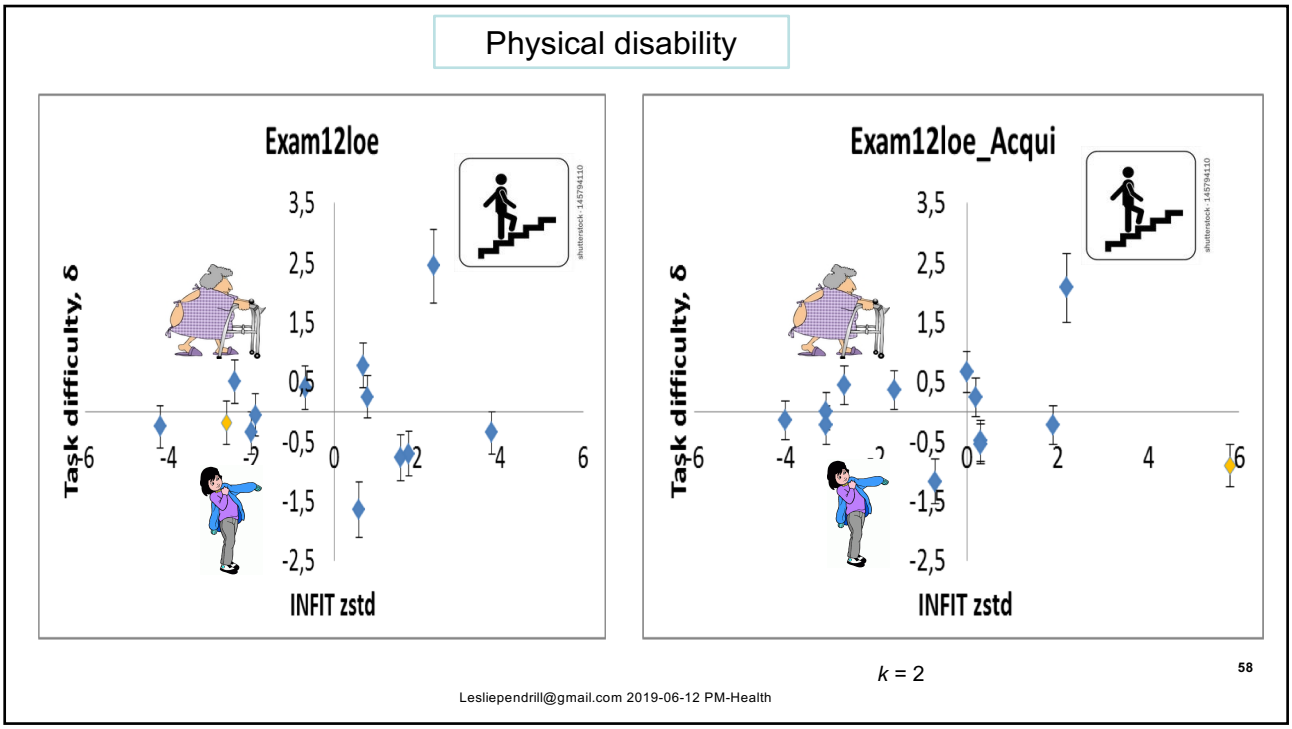


DM1-Activ^{C} scale of 105 items

Myotonic dystrophy type 1

56


Physical ergonomics
Construct Alley - Activity
Emotion
Cognitive
Motor
More challenge
Less challenge
δ (logit)
6
4
2
0
-2
-4
-10
-5
0
5
Grooming: Care for one's personal appearance, hygiene, and clothing
carry/put down a heavy object
lift heavy object (10 kilograms)
walk up 3 stairs
walk uphill
walk in the dunes
walk upstairs with a bag
carry shopping
walk on uneven ground
mop the floor
walk up 2 stairs
clean the toilet
bend and pickup
button shirt/blouse
put away the dishes
put on shoes
dry your body
take a shower
dress your lower body
hang up a coat
put on a coat
dress your upper body
put on a sweater
put on a t-shirt
make coffee/tea
clip your toe nails
stand up from lying
clip your finger nails
Infit_zstd_WH
peel an apple/orange
lie down from standing
pick up a small object
fill in a form
handle small objects
work on a computer
sit down from standing
read a book
read a newspaper
peel a banana
Myotonic dystrophy type 1
DM1-Activ^C scale of 105 items


Physical disability
Exam12loe
Exam12loe_Acqui
Taşk difficulty, δ
INFIT zstd
3,5
2,5
1,5
0,5
-0,5
-1,5
-2,5
-6
-4
-2
0
2
4
6
k = 2
Lesliependrill@gmail.com 2019-06-12 PM-Health



	A	B	C	D		F	G	H	I		K	L	M
1	TP	A. EATING	B. GROOMING	C. BATHING	D. UPPER BODY DRESSING	E. LOWER BODY DRESSING	F. TOILETING	G. BLADDER	H. BOWEL	I. BED TRANSFER	J. TOILET TRANSFER	K. TUB, SHOWER	L. WALK/WHEELCHAIR	M. STAIRS
2	21101	5	5	2	3	1	3	3	3	2	2	1	2	1
3	21170	4	4	3	3	4	4	3	3	4	5	4	5	4
4	21174	4	4	4	4	3	4	4	4	3	4	1	4	1
5	21191	4	4	4	4	2	4	4	4	4	4	3	4	4
6	21243	4	2	2	1	1	2	2	2	1	1	1	1	1
7	21250	3	2	1	2	2	2	4	4	4	4	4	5	1
8	21266	4	4	3	4	2	3	5	5	4	4	3	2	1

DIF – differential item functioning

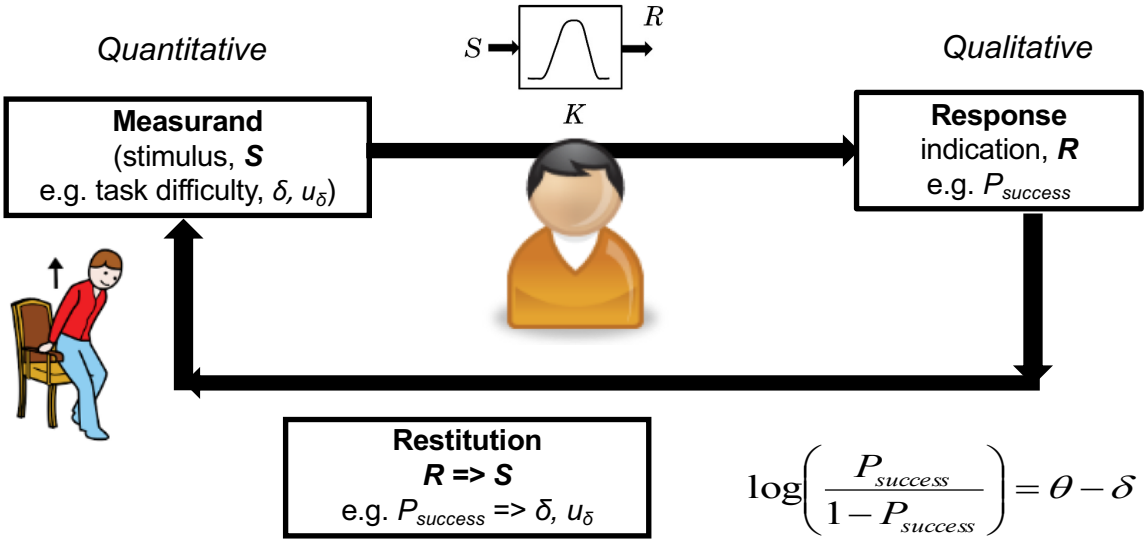
”The symptom of DIF is that a "focal" group does better (or worse) on a particular item than a "reference" group, given those groups' performance on the rest of the test.”

Y Du, F Yates 1995 “When to adjust for Differential Item Functioning”, *Rasch Measurement Transactions*, **9**:1, p.414, https://www.rasch.org/rmt/rmt91e.htm

59

Lesliependrill@gmail.com 2019-06-12 PM-Health

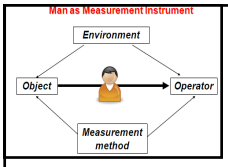
Measurement system analysis



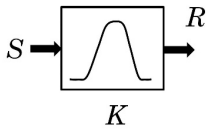
L R Pendrill 2018 *Meas. Sci. Technol.* https://doi.org/10.1088/1361-6501/aa9cd2

60

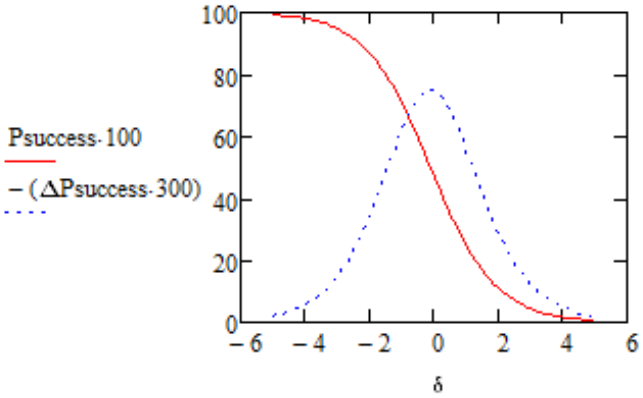
Lesliependrill@gmail.com 2019-06-12 PM-Health



Sensitivity, *K*

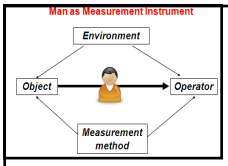


$$K = \frac{\partial P_{success}}{\partial \delta} = \frac{e^{2\cdot(\theta-\delta)}}{(1+e^{(\theta-\delta)})^2} - \frac{e^{(\theta-\delta)}}{1+e^{(\theta-\delta)}}$$



$$P_{success} = \frac{e^{\theta-\delta}}{1+e^{\theta-\delta}}$$

61



Object value

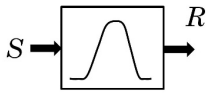
Task difficulty, δ

$\delta = \delta_0$

Instrument

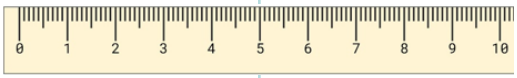
Decision risks

$$\mathbf{P} = \begin{pmatrix} 1-\alpha & \alpha \\ \beta & 1-\beta \end{pmatrix}$$

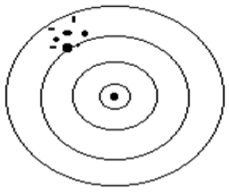


Instrument (person) ability, θ

Response



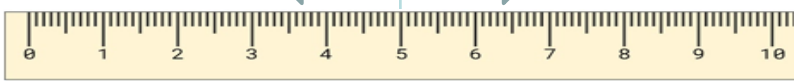
Bias $\delta = \delta_0 + \varepsilon_\delta$



Low trueness
High precision

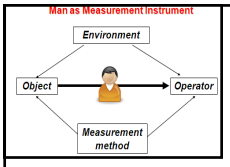
$\Delta\delta = k \cdot \delta_0$

Re-scaling



$$P_{success} = 1-\alpha = \frac{e^{\theta-\delta}}{1+e^{\theta-\delta}}$$

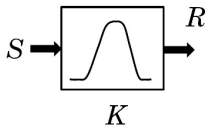
62



Response

$$P_{success} = \frac{e^{\theta-\delta}}{1+e^{\theta-\delta}}$$

Sensitivity



Fit residual $y_{i,j} = x_{i,j} - \mathbb{E}_{i,j}$ $\quad {\chi_j}^2 = \sum_{i=1}^{N_{TP}} {y_{i,j}}^2$

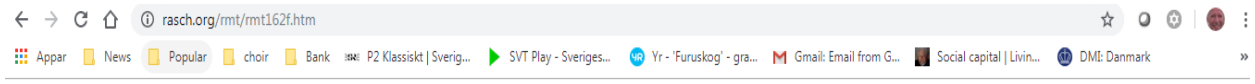
$$K = \frac{\partial P_{success}}{\partial \delta} = \frac{e^{2\cdot(\theta-\delta)}}{(1+e^{(\theta-\delta)})^2} - \frac{e^{(\theta-\delta)}}{1+e^{(\theta-\delta)}}$$

$$\chi'_j = \chi_j + \frac{\partial P_{success}}{\partial \theta}\cdot\partial\theta + \frac{\partial^2 P_{success}}{\partial \theta^2}\cdot\theta$$

$$\chi'_j = \chi_j - \frac{\partial P_{success}}{\partial \delta}\cdot\partial\delta - \frac{\partial^2 P_{success}}{\partial \delta^2}\cdot\delta$$

Changed stimulus

Changed sensitivity



What do Infit and Outfit, Mean-square and Standardized mean?

These are all "fit" statistics. In a Rasch context they indicate how accurately or predictably data fit the model. Dichotomous fit statistics. Polytomous fit statistics.

Infit means inlier-sensitive or information-weighted fit. This is more sensitive to the pattern of responses to items targeted on the person, and vice-versa. For example, infit reports overfit for Guttman patterns, underfit for alternative curricula or idiosyncratic clinical groups. These patterns can be hard to diagnose and remedy.

What do Infit and Outfit, Mean-square and Standardized mean? Linacre JM. … Rasch Measurement Transactions, 2002, 16:2 p.878

Molton: "Both statistics … require a lot of meditation and incense. For useful explanations, refer to: http://www.rasch.org/rmt/rmt162f.htm".

THE DISTRIBUTION OF CHI-SQUARE

By Edwin B. Wilson and Margaret M. Hilferty

Department of Vital Statistics, Harvard School of Public Health

Communicated November 6, 1931

R. A. Fisher[1] gives a table of χ^2 and states that for large values of n, the number of degrees of freedom in the distribution,

$\sqrt{2\chi^2} - \sqrt{2n-1}$ is normally distributed with $\sigma = 1$. (1)

It is interesting to ask what other formulas of a similar sort might be used.

$\left(\frac{\chi^2}{n}\right)^{1/3}$ is normally distributed about $1 - \frac{2}{9n}$ with $\sigma^2 = \frac{2}{9n}$. (5)

$\sqrt[3]{\chi^2}$ about $\sqrt[3]{n-2/3}$ with $\sigma = \frac{\sqrt{2}}{3\sqrt[6]{n-2/3}}$. (4)

$$Chi^2 = \frac{X - N_{TP}}{\sqrt{2 \cdot N_{TP}} \cdot \sigma}$$

$$Fisher = \frac{\sqrt{2 \cdot X} - \sqrt{2 \cdot N_{TP} - 1}}{\sigma}$$

$$WH = \frac{2 \cdot \left(\sqrt[3]{X} - \sqrt[3]{N_{TP} - \frac{2}{3}}\right)}{\sigma}$$

$$\sigma = \frac{\sqrt{2}}{3 \cdot \sqrt[6]{N_{TP} - \frac{2}{3}}}$$

65

Score residual, *y*

$$y_{i,j} = x_{i,j} - \mathbb{E}_{i,j}$$

Expected mean, *E*, of response *x*

$$\mathbb{E}_{i,j} = \sum_{k=0}^{C_j} k \cdot q_{i,j,k}$$

$$q_{i,j,k} = \frac{e^{\sum_{c=0}^{k}(\theta_i - \delta_{j,c})}}{\sum_{k=0}^{C_j} e^{\sum_{c=0}^{k}(\theta_i - \delta_{j,c})}}$$

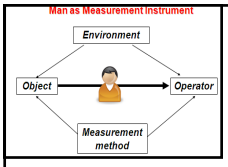
$$X = {\chi_j}^2 = \sum_{i=1}^{N_{TP}} {y_{i,j}}^2$$

$$INFIT\_zstd\_WH = \frac{2 \cdot \left(\sqrt[3]{X} - \sqrt[3]{N_{TP} - \frac{2}{3}}\right)}{\sigma}$$

$$\sigma = \frac{\sqrt{2}}{3 \cdot \sqrt[6]{N_{TP} - \frac{2}{3}}}$$

Molton: “Both statistics … require a lot of meditation and incense. For useful explanations, refer to: http://www.rasch.org/rmt/rmt162f.htm”.

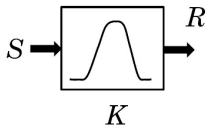
66



Response

$$P_{success} = \frac{e^{\theta-\delta}}{1+e^{\theta-\delta}}$$

Sensitivity



$$K = \frac{\partial P_{success}}{\partial \delta} = \frac{e^{2\cdot(\theta-\delta)}}{(1+e^{(\theta-\delta)})^2} - \frac{e^{(\theta-\delta)}}{1+e^{(\theta-\delta)}}$$

Fit residual $y_{i,j} = x_{i,j} - \mathbb{E}_{i,j}$

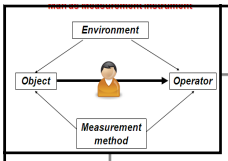
$$X = \chi_j{}^2 = \sum_{i=1}^{N_{TP}} y_{i,j}{}^2$$

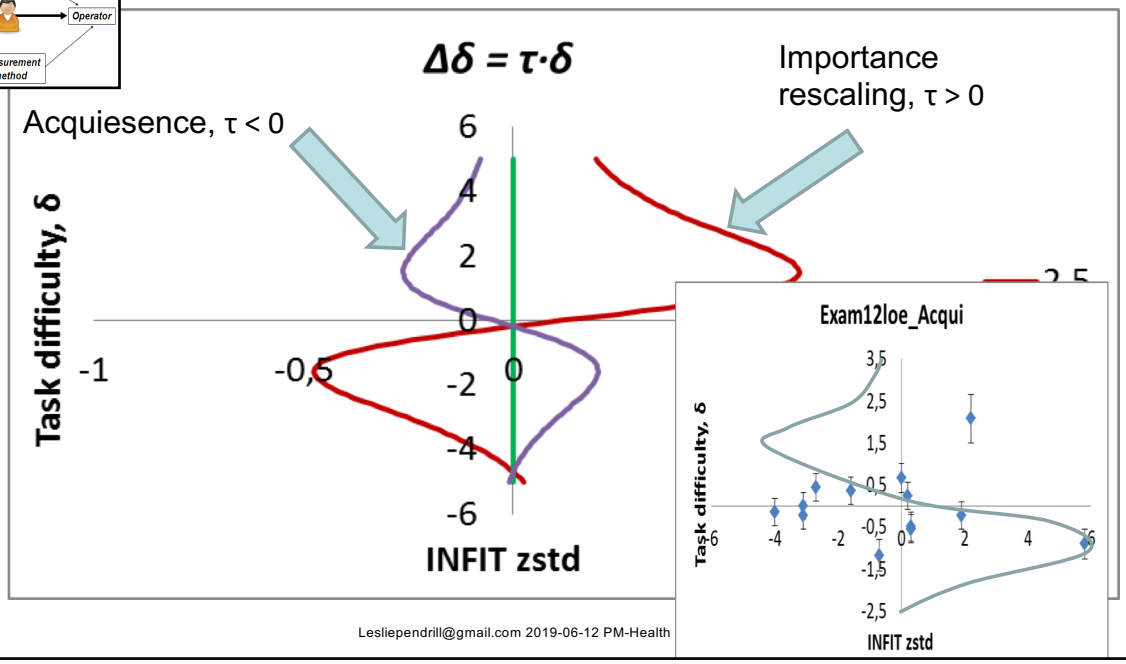
$$\chi'_{j,\delta}{}^2 = \left(\chi_j - \frac{\partial P_{success}}{\partial \delta}\cdot\partial\delta\right)^2 = \chi_j{}^2 - 2\cdot\frac{\partial P_{success}}{\partial \delta}\cdot\partial\delta\cdot\chi_j + \left(\frac{\partial P_{success}}{\partial \delta}\cdot\partial\delta\right)^2$$

Changed stimulus

$$Infit\_zstd\_WH = \frac{2\cdot\left(\sqrt[3]{X} - \sqrt[3]{N_{TP} - \frac{2}{3}}\right)}{\sigma}$$

67





68

Restitution in case of performance metric response

Response, *y*,

from object *A*
(attributes θ, δ)

$$P(y|a); \varphi = P_{success} = \frac{e^{(\theta-\delta)}}{1+e^{(\theta-\delta)}}$$

Restituted entity quantity, z_R,

of object *A*
(attributes θ, δ)

$$z_R = S = \theta - \delta = log\left[\frac{P_{success}}{1-P_{success}}\right]$$

69

Restitution in case of performance metric response

Response, *y*,

from object *A*
(attributes θ, δ)

via instrument
(attributes *b*, *ρ*)

$$P(y|a); \varphi = P_{success} = b + (1-b) \cdot \frac{e^{\rho\cdot(\theta-\delta)}}{1+e^{\rho\cdot(\theta-\delta)}}$$

3PL

Restituted entity quantity, z_R,

of object *A*
(attributes θ, δ)

via instrument
(attributes *b*, *ρ*)

$$z_R = S = \theta - \delta = log\left[\frac{\frac{P_{success}}{1-b} - b}{1 - \frac{P_{success}}{1-b} - b}\right] - log(\rho)$$

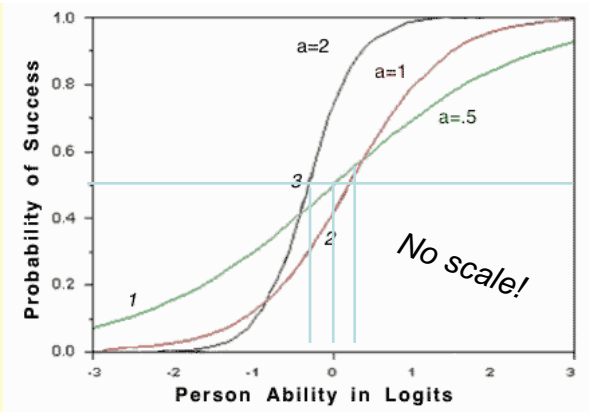
70

Rasch v 2PL

Humphry and Andrich [2008]:

- Incorporation in an Item Response Theory model of a discrimination parameter which is estimated for each item (or person)
- will in general break conditions for sufficiency and specific objectivity,
- and thus the opportunity of establishing units and measurement scales.

http://hosted.jalt.org/test/sic_5.htm



3 1 2

Discrimination, ρ

$$2PL\_P_{success} = \frac{e^{\rho\cdot(\theta-\delta)}}{1+e^{\rho\cdot(\theta-\delta)}}$$

71

Logistic measurement function

Conditions for sufficiency and specific objectivity, and thus opportunity of establishing units and measurement scales

- maintained if one associates a discrimination factor (ρ_s) with a set, s, of items rather than a single item

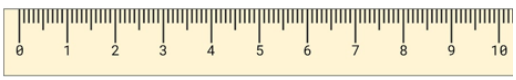
S M Humphry 2011, "The Role of the Unit in Physics and Psychometrics", *Measurement: Interdisciplinary Research and Perspectives*, **9**(1): p. 1-24

$$ln\left(\frac{P_{success,i,j}}{1-P_{success,i,j}}\right) = \rho_s \cdot \left(\theta_i^* - \delta_j^*\right)$$

Modified Rasch parameters $\theta_i^* = \frac{\theta_i}{\rho}$ $\delta_j^* = \frac{\delta_j}{\rho}$

Measurement units: $\theta_i = \{\theta_i\}\cdot[\theta]$ $\delta_j = \{\delta_j\}\cdot[\delta]$

'Common unit' of measure: $[\theta] = [\delta]$



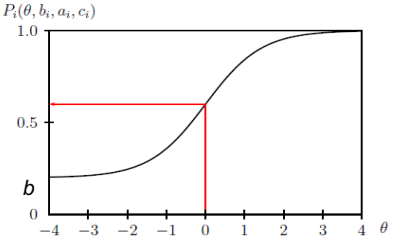
72

A Critique of 3-PL IRT Estimation

Bias, b

Discrimination, ρ

$$3PL\_P_{success} = b + (1 - b) \cdot \frac{e^{\rho \cdot (\theta - \delta)}}{1 + e^{\rho \cdot (\theta - \delta)}}$$



b = probability of correct response when true ability $\theta \rightarrow -\infty$

Number of response categories for item i, C_i

'Guessing' probability, $\pi_{i,j}$

$$3PL\_P_{success,i,j} = \pi_{i,j} \cdot \frac{1}{C_i} + (1 - \pi_{i,j}) \cdot \frac{e^{\rho_i \cdot (\theta_i - \delta_j)}}{1 + e^{\rho_i \cdot (\theta_i - \delta_j)}}$$

73

https://www.rasch.org/rmt/rmt272a.htm

3PL IRT or Rasch?

Wright BD 1995 Rasch Measurement Transactions, 9:1 p.408

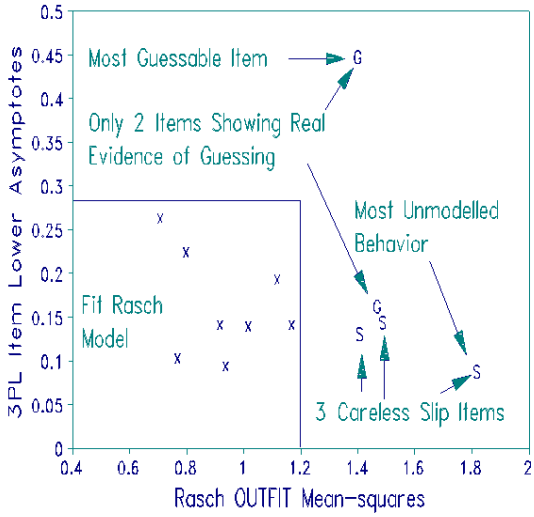
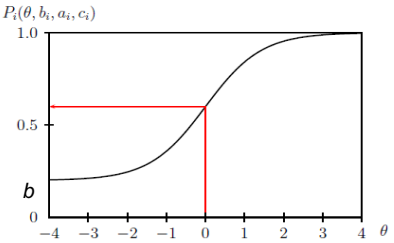


Figure 3: Guessability and OUTFIT mean-squares



3PL:

Number of response categories for item i, C_i

'Guessing' probability, $\pi_{i,j}$

$$3PL\_P_{success,i,j} = \pi_{i,j} \cdot \frac{1}{C_i} + (1 - \pi_{i,j}) \cdot \frac{e^{\rho_i \cdot (\theta_i - \delta_j)}}{1 + e^{\rho_i \cdot (\theta_i - \delta_j)}}$$

74

A Critique of 3-PL IRT Estimation

1. Mathematical analysis shows that the 3p model is a non-converging, inestimable elaboration of the Rasch model.

3. Whenever something must be done about the few lucky guesses which actually occur in multiple choice item response data, the few persons responsible for those occurrences are easy to find and reasonable corrections for any interference with measurement are easy to apply (Wright & Stone, 1979, pp. 170-190).

Number of response categories for item i, C_i

'Guessing' probability, $\pi_{i,j}$

$$3PL\_P_{success,i,j} = \pi_{i,j} \cdot \frac{1}{C_i} + \left(1 - \pi_{i,j}\right) \cdot \frac{e^{\rho_i \cdot (\theta_i - \delta_j)}}{1 + e^{\rho_i \cdot (\theta_i - \delta_j)}}$$

Benjamin Drake Wright, 12/18/95, in a Note to Allan Olson, Northwest Evaluation Association (NWEA).

https://www.rasch.org/rmt/rmt272a.htm

A Critique of 3-PL IRT Estimation

The 3p model buries this individual person information by forcing item guessing parameters on everyone who takes the items whether they guess or not.

If something beneficial, not to mention legal, is to be done about guessing, then it must face those few persons who benefit from lucky guesses and not mistreat everyone else.

Number of response categories for item i, C_i

'Guessing' probability, $\pi_{i,j}$

$$3PL\_P_{success,i,j} = \pi_{i,j} \cdot \frac{1}{C_i} + \left(1 - \pi_{i,j}\right) \cdot \frac{e^{\rho_i \cdot (\theta_i - \delta_j)}}{1 + e^{\rho_i \cdot (\theta_i - \delta_j)}}$$

Benjamin Drake Wright, 12/18/95, in a Note to Allan Olson, Northwest Evaluation Association (NWEA).

https://www.rasch.org/rmt/rmt272a.htm

Lesliependrill@gmail.com 2019-06-12 PM-Health

IRT or Rasch – what can and cannot be done?

- Man as Measurement Instrument
- Quality-assured measurement
- Logistic ruling & counted fractions
- Uncertainty & bias in measurement systems
- Classical test theory or Rasch?
- Aquiescence and Construct Alleys
- Rasch or 3PL IRT?

77

Acknowledgments

Work performed as part of 15HLT04 NeuroMet project, European Metrology Programme for Innovation & Research (Horizon 2020), jointly funded by EMRP participating countries within EURAMET (www.euramet.org) and European Union.

Thanks are due, particularly, to:
- Members of the EMPIR NeuroMet consortium
- Jeanette Melin, RISE Metrology Göteborg Sweden
- Stefan Cano, Modus Outcomes, Boston (USA) & Stotford UK
- William P. Fisher, Jr., Ph.D., Research Associate, BEAR Center, Graduate School of Education, University of California, Berkeley, CA (USA) & Principal, LivingCapitalMetrics Consulting

78

Acknowledgments

NeuroMET : www.lgcgroup.com/EMPIR-neuromet

This project has received funding from the EMPIR programme co-financed by the Participating States and from the European Union's Horizon 2020 research and innovation programme

NeuroMET: *Innovative measurements for improved diagnosis and management of neurodegenerative diseases*

EURAMET's Research Programmes

- EURAMET's research programmes (EMRP and EMPIR) support the collaboration of European metrology institutes, industrial organisations and academia through Joint Research Projects (JRPs). They are structured around European Grand Challenges in such areas as Health, Energy, the Environment and also aim to progress fundamental science.
- See https://www.euramet.org/ for more details.

EMRP
European Metrology Research Programme
Programme of EURAMET
The EMRP is jointly funded by the EMRP participating countries within EURAMET and the European Union

EMPIR EURAMET
The EMPIR initiative is co-funded by the European Union's Horizon 2020 research and innovation programme and the EMPIR Participating States

80

Lesliependrill@gmail.com 2019-06-12 PM-Health