PMhealth – Nationell workshop i Psykometri och Metrologi inom hälsovetenskaperna 2019

Högskolan Kristianstad
11 – 13 June, 2019
https://www.hkr.se/pmhealth2019

WIRTSCHAFTS UNIVERSITÄT WIEN VIENNA UNIVERSITY OF ECONOMICS AND BUSINESS

Six decades of measurement using the Rasch model in the social sciences – Setting the agenda for the next sixty years

(and a coffee break)

("The future glimmers long before it comes to be")

JOINT SYMPOSIUM IMEKO2019 TC1 TC7 TC13 TC18

Thomas Salzberger
thomas.salzberger@wu.ac.at
University of Economics and Business, Vienna, Austria (WU WIEN)

Personal Background

Why Measurement in the Social Sciences?

Numbers and Measurement – The Role of the Measurement Model
A Very Simple Approach

Requirements of Measurement

Does the Very Simple Approach Deliver?

Is Measurement in the Social Sciences Different from Measurement in the Natural Sciences?

A not Quite so Simple Approach

Comparing CTT, IRT and the Rasch Model/RMT

A case for the Rasch Model/RMT

The Rasch Model: the first 60 years

The Rasch Model: the next 60 years

The Rasch Model and Metrology

Revisiting Feynman

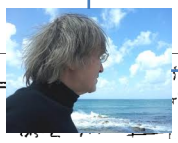
A G E N D A

3

Personal Background

Studies of business administration with focus on marketing (research)

Studies of psychology



Georg Rasch (1901-1980)

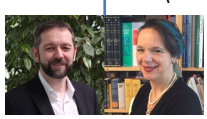
Reinhold Hatzinger (1953-2012)

Regina Dittrich

*Klaus Kubinger (*1949)*

*Gerhard Fischer (*1938)*

Anton Formann (1950-2010)



Ivo Ponocny and Elisabeth Ponocny-Seliger

The Rasch model for measurement

Georg Rasch's blackboard in Copenhagen (negative by Michele Bruniges)

LpcM-Win

Structural Rasch modeling. Handbook of the usage of LpcM-Win 1.0.

4

Personal Background

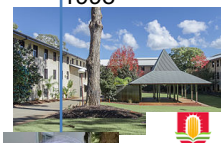
Studies of business administration with focus on marketing (research)

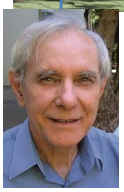
*Hartmut Holzmüller (*1955)*

Studies of psychology

Georg Rasch (1901-1980)

1998





Murdoch UNIVERSITY

*David Andrich (*1941)*

Doctoral thesis on measurement equivalence based on CFA and RM (1994-1998)

Habilitation on Measurement in Marketing Research An alternative framework (2003-2008/09)

Special role of the Rasch model within Item Response Theory

5

Personal Background

Measurement in Marketing

Measurement in the Social Science

Measurement in Health

Measurement in Science

ERRTG European Rasch Research and Teaching Group

Stefan Cano

William Fisher and Leslie Pendrill

Joel Michell

and many many more ...

6

Why Measurement?

Quantitative approach to science proved enormously successful in the natural sciences.

"The world is built upon the power of numbers."

Pythagoras

"All is number."

Explaining the world using mathematics.

Mathematics: "that which is learned"

7

Why Measurement in the Social Sciences?

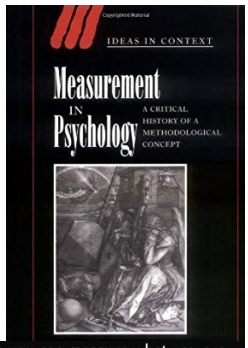
Quantitative science has become the role model for the social sciences

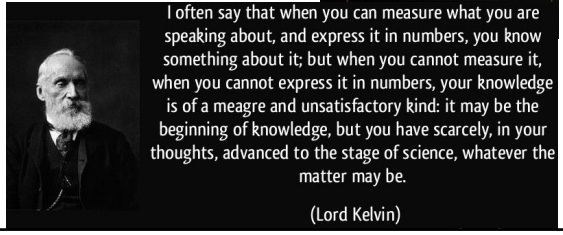
Quantitative imperative (Joel Michell)

"Measure what can be measured, and make measurable what cannot be measured"

- Galileo Galilei (perhaps)

"Two things are infinite: the universe and human ingenuity to make up quotes; and I'm not sure about the universe."
Max Planck





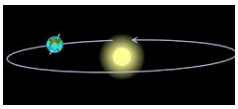
What Does Measurement Provide?

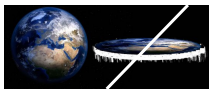
Measurement is an important **epistemological window** into reality

Effectiveness and trustworthiness of **quantitative science** hinges on proper measurement *and* substantive theory

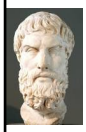
9

Today, the Earth Orbits the Sun. It wasn't always so. Or was it?





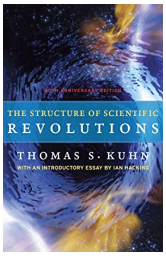
Roman Catholic doctrine:
Earth at the centre of the universe
Echoed Greek geographer Ptolemy (2nd century AD)



Even though Aristarchus of Samos (3rd/2nd century BC) proposed that the Earth orbits the Sun

But no empirical evidence was available
Theory on its own is pointless

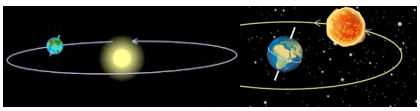
Heliocentric theory only revived in the 16th century by Copernicus
New empirical evidence: Mathematics & Observations confirmed heliocentric theory



And yet did Copernicus dare to publish his results only in 1543 - the year of his death

10

Today, the Earth Orbits the Sun. It wasn't always so. Or was it?



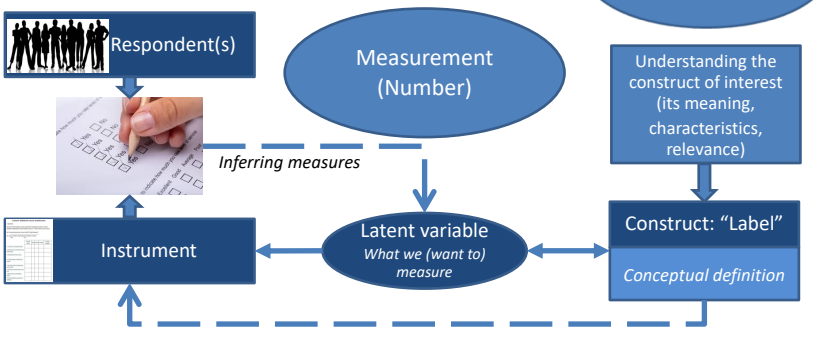
Galilei collected further evidence of heliocentric theory
Galilei's contribution (among others) was the use of mathematics in physics

Withdrew due to massive resistance from the Roman Catholic Church (think of different paradigms!)

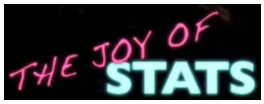
Measurement as quantification involves **numbers**
The reverse is not necessarily true …

11

How Do We Get the Numbers in the Social Sciences?



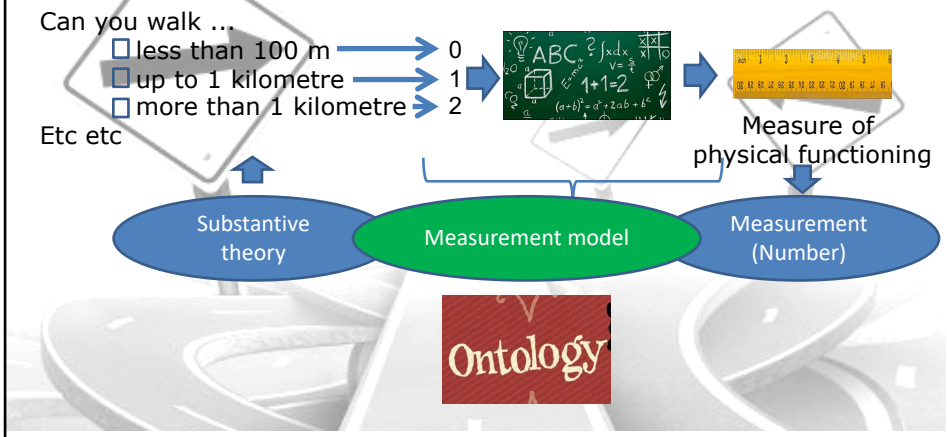
How exactly do we infer numbers?
We need a measurement model.



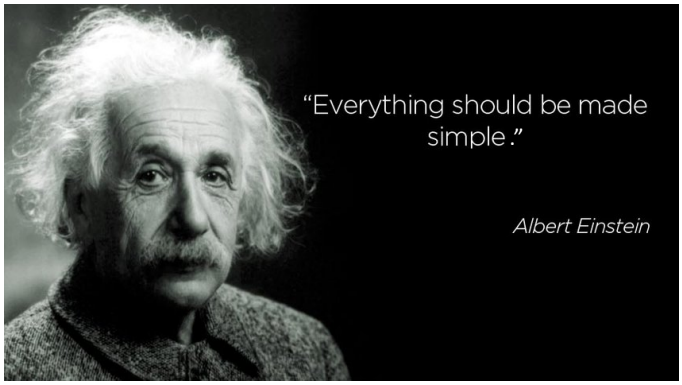
12

Measurement Models

Manifest response (qualitative) -> transformed into count by scoring (ordinal) -> inference of measures by measurement model (number)



13



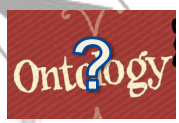
14

Measurement Models

Manifest response (qualitative) -> transformed into count by scoring (ordinal) -> inference of measures by measurement model (number)

S.S. Stevens

Measurement is the assignment of numerals to objects according to rules



Processing presumed measures

Can you walk ...

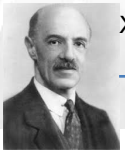
- ☐ less than 100 m → 0
- ☐ up to 1 kilometre → 1
- ☐ more than 1 kilometre → 2

Classical Test Theory (CTT) (True Score Theory) → Overall measure

$X = T + E$

Factor analysis

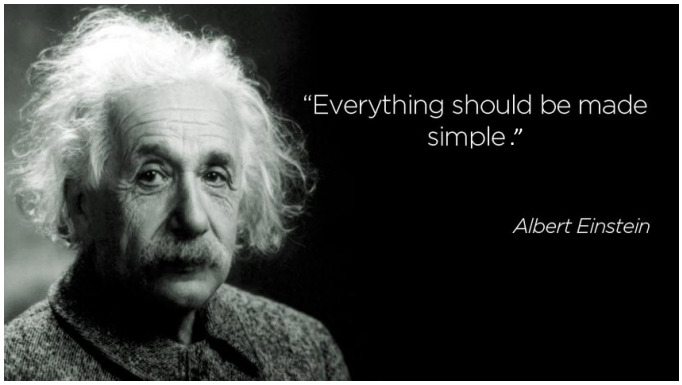
$x_{iv} = \tau_i + \lambda_i \xi_v + \delta_{iv}$



C. Spearman

L.L. Thurstone

15



16

Numbers and Measures

What we end up with:
a number

... which can be many different things

A proper measure,
An amount expressed in an uncertain or varying metric,
A count,
A summary of measures for some purpose,
Merely a numeral disguised as a measure

What we need:
a measure

Justification of the scoring (raw score sufficiency),
Invariant ruler (invariance, specific objectivity),
Stable metric (unit of measurement),
Measure as an interpretable amount of the property measured (link to substantive theory),
Statement of uncertainty (measurement error),
Transcending dependency on specific instrument

Treating numbers as measures, requires a strong justification

17

Measurement and Sufficiency

Measurement is an exhaustive, unambiguous, parsimonious and meaningful description of a respondent

We cannot know more of the person with respect to the attribute measured

Combining observations (adding up ratings to a score) has to be based on a minimally sufficient statistic

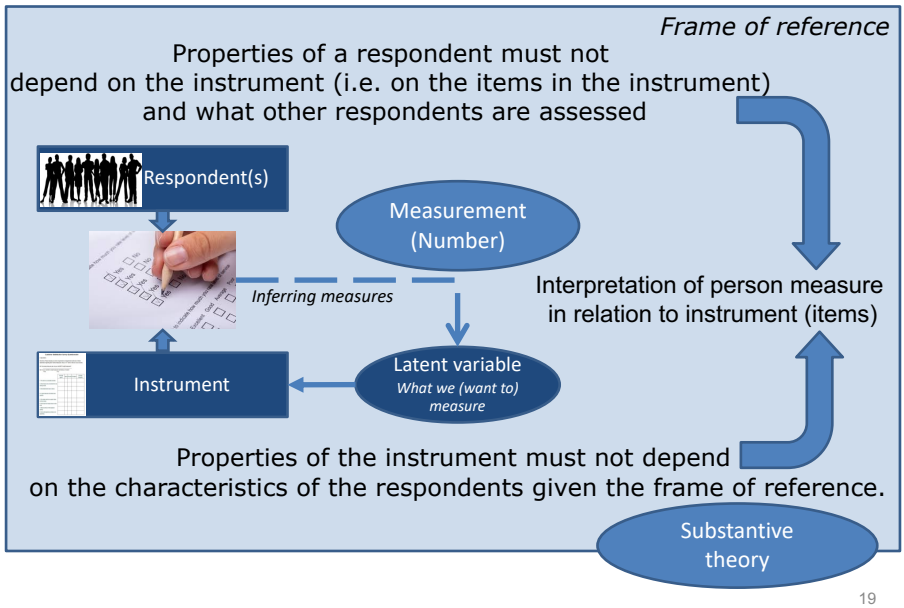
Raw score sufficiency is not merely a statistical convenience but a requirement for measurement

Summary scores (indexes) that combine multiple attributes and describe them by one number

- have no unambiguous meaning
- have a different ontological status (social construction)

18

Measurement, Invariance and the Frame of Reference



19

Measurement and the Unit of Measurement

Measures must have a stable unit of measurement (linearity)
Raw scores are non-linear

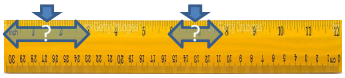
Unit of measurement must be invariant
Measurement based on models assuming a particular person distribution (typically normal) lacks an invariant unit

Unit of measurement must to be interpretable
Factor scores merely express percentile ranking at best

Unit of measurement ought to transcend a particular instrument
Requires theory-based unit of measurement

20

Measurement and Uncertainty



Uncertainty of measurement must be invariant with respect to the sample of persons

Precision of a person measurement must not depend on who else has been assessed

Uncertainty of measurement must reflect actual precision of the measurement instrument along the measurement continuum

One and the same standard of error across the whole range is implausible

Paradox of consistency of extreme scores

21

Does CTT/TTT Deliver Measurement?

Scoring

First of all, we have to abandon Stevens' definition of measurement. Otherwise everything becomes tautological.

CTT does not justify the scoring.
In factor analysis, the weighting of item scores is sample-dependent, hence no invariance.

Invariance

CTT assumes a normal distribution of persons.
Reliance on inter-item-correlations, which are sample-dependent.

Stable unit of measurement

CTT relates a non-linear (at best) score by a linear transformation to a measure, which therefore remains non-linear.

22

Does CTT/TTT Deliver Measurement?

Interpretable measures

CTT raw score measures are non-linear and remain unjustified. Factor scores related to person ranking.
Strong focus on item-interrelationship, which is almost impossible to predict based on theory or explain meaningfully post-hoc. No item hierarchy.

Uncertainty

Measurement precision ("reliability") stated at the instrument level but dependent on distributional assumptions and the sample.
Standard error derived from reliability applied to any level of the measurement continuum.

Transcending dependency on specific instrument and instance of application

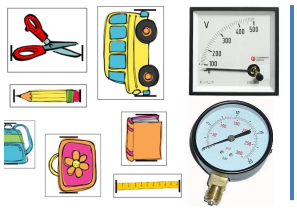
Given the limitations of CTT measures, attempts are questionable to begin with.
Equating remains sample dependent.

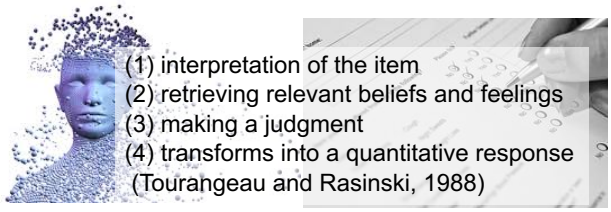
23

But Maybe Measurement in the Social Sciences is Different Altogether

Measurement in the social sciences certainly differs in many ways from measurement in the natural sciences.

- Interaction of a very special object (subject capable of cognitions and meta-cognitions) and an instrument (see Pendrill, 2014)
- Replicability of measurement, etc.





But measurement does not differ in terms of the intended nature of the outcome and its interpretation

One number that exhaustively and unambiguously represents the amount of the property the person possesses

24

Measurement in the Social Sciences and the Natural Sciences

Epistemological difference
- *How we observe*
- *How we process observations*
- *Different "measurement machinery"*

Same ontological claim

Physical measurement	Social measurement
Properties exist independently of their being measured	Properties exist independently of their being measured
Measurement does not create properties	Measurement does not create properties
Properties as a function of fundamental laws of nature	(At least some) properties are a function of society
Wide frame of reference (universe, the world we live in), but possibly not unbounded	Narrow frame of reference (e.g., properties may be limited to a particular society, or pertain to time)

25



The Rasch Model as an Alternative to CTT/TTT (& IRT)

The Rasch Model for Measurement
For dichotomous data
(Rasch, 1960)

$$P(a_{vi} = 1) = \frac{e^{\beta_v - \delta_i}}{1 + e^{\beta_v - \delta_i}}$$

For polytomous data (ordered categorical data)
(Andrich, 1978, 1988)

$$P(a_{vi} = x \mid \beta_v, \tau_{ij}, j = 1 \ldots m, 0 < x \leq m) = \frac{e^{\left(\sum_{j=1}^{x} -\tau_{ij}\right) + x \cdot (\beta_v - \delta_i)}}{\Upsilon}$$

By its form, the Rasch model is an Item Response theory (IRT) model
- But fundamentally different philosophy underpinning non-Rasch IRT and the Rasch Model/RMT

26

The Rasch Model & IRT

$$P(a_{vi} = 1) = \frac{e^{\beta_v - \delta_i}}{1 + e^{\beta_v - \delta_i}}$$

Rasch Model

$$P_i(X_i = 1 \mid \theta_j) = \frac{\exp[Da_i(\theta_j - b_i)]}{1 + \exp[Da_i(\theta_j - b_i)]}$$

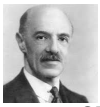
Birnbaum Model (2pl)
(Graded Response Model)


Allan Birnbaum & Frederic Lord

Based on fundamental principles of measurement	Aiming at best accounting for the data based on some assumptions of the respondent distribution
Prescriptive model	Descriptive model
Invariance (specific objectivity)	No invariance

27

The Rasch Model, CTT and IRT




CTT

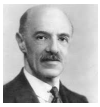
IRT (incl. 1pl)

RMT

	CTT	IRT (incl. 1pl)	RMT
Scoring/ Invariance	Unweighted item score not justified, interpreted as measure; weighted factor score (FA) sample dependent; widespread habit of using the raw score as a measure essentially implies data fitting the RM	Unweighted score is <u>not</u> a sufficient statistic, joint estimation of item and person properties based on distributional assumptions; meaningfulness hinges on the model fitting the data and data meeting the distributional assumptions; no unequivocal transformation of unweighted score to a measure (despite many papers claiming so and implicitly assuming a RM)	Unweighted score is a sufficient statistic, interpreted as a summary of what is observed, meaningfulness hinges on data fitting the model, justification of order can be tested; invariant measures based on an unequivocal score to measure transformation

28

The Rasch Model, CTT and IRT




	CTT	IRT (incl. 1pl)	RMT
Unit of measure-ment/ Inter-pretable measures	non-linear unit at best (speculative); population-related inter-pretation of measure	Linear unit but speculative, depends on data meeting the distributional assumptions, and the description of unequal item unit (by means of estimating item discrimination) being true (assuming a frame-of-reference RM holds) measure population-related, interpretation complicated by the item hierarchy varying	Linear unit of measurement; person measures interpretable in terms of unequivocal hierarchy of items

29

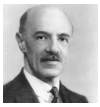
The Rasch Model, CTT and IRT




	CTT	IRT (incl. 1pl)	RMT
Un-certainty	concept of reliability confounds population characteristics and precision; S.E.M. population/ sample-dependent	S.E.M. depends on available information, which, however, is population-dependent	S.E.M. depends on available information (the items actually administered), information is population-independent by virtue of invariance

30

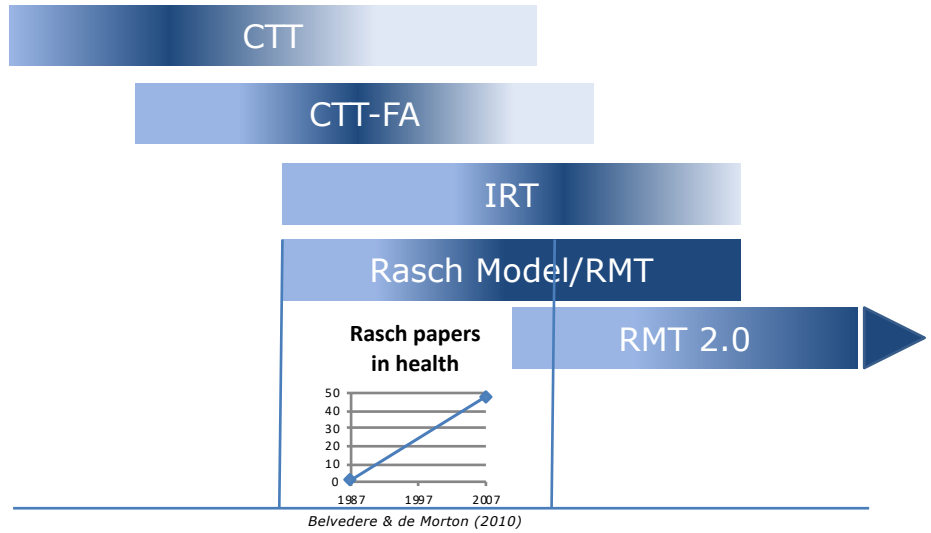
The Rasch Model, CTT and IRT




	CTT	IRT (incl. 1pl)	RMT
Tran-scending instru-mentation and instance of application	linking instruments by equipercentile equating, sample/population-dependent	sample/ population-dependent	establishing a common metric by co-calibration of instruments, sample-independent

31

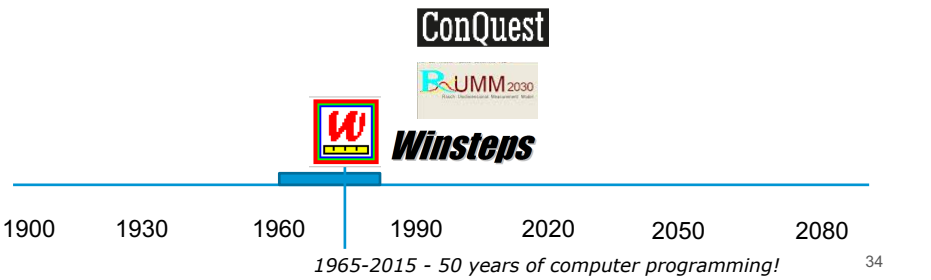
The First 60 Years of the Rasch Model



33

Fundamental Contributions

- The Rasch model for dichotomous responses (Rasch, 1960, 1961)
- Specific objectivity (Rasch, 1977)
- Justifying polytomous item scoring (Andersen, 1977; Andrich, 1978)
- The Rasch model for polytomous responses (Andrich, 1978; Masters, 1982)



1965-2015 - 50 years of computer programming!

34

Recent Contributions

Understanding the Rasch Measurement Model

Formal requirements, how the model works, understanding and handling of problems

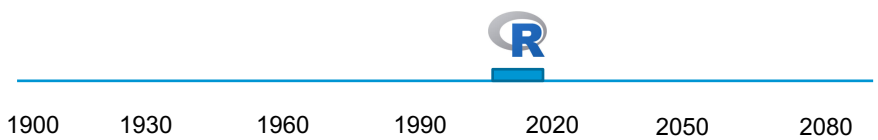
- *Longitudinal measurement (Olsbjerg & Christensen, 2015; Andrich, 2016).*
- *Local dependence (Marais & Andrich, 2008; Andrich & Kreiner, 2010; Andrich et al., 2012).*
- *Frame of reference: impact on unit (Humphry, 2005; Humphry & Andrich, 2008)*
- *Differentiating true and artificial differential item functioning (Andrich & Hagquist, 2012, 2015; Hagquist & Andrich, 2015)*
- *Scoring polytomous items/threshold structure (Andrich, 2013)*
- *Guessing in multiple choice items (Andrich et al., 2012; Andrich & Marais, 2014)*

35

Recent Contributions

Understanding fit statistics

- *Tests of fit and their interpretation (Kubinger et al., 2011)*
- *Insight into distributional properties (Christensen et al., 2017; Marais, 2013; Müller, 2015, 2018)*
- *Insight into the role of the sample size (Hagell & Westergren, 2016; Draxler & Alexandrowicz, 2015; Draxler, 2010; Hobart et al., 2012; Kubinger et al., 2009)*



36

Psychometric Justification of the Rasch Model

Theory → Data

Rasch Model embodies necessary requirements of measurement in the social sciences

- Statistical measurement model: Formal theory of measurement
- Suitable to falsify measurement attempt due to its rigorous prescriptive character
- Provides measures expressed in a stable unit of measurement (implicit unit)
- Provides estimate of uncertainty
- Provides strong evidence of validity

Rasch Measurement Theory
(Andrich, 2017; Ewing, Salzberger, Sinkovics, 2005)

37

Understanding the Concept to be Measured

Data → Conceptual Theory

- Understanding of the characteristics of the attribute measured
- Understanding of the hierarchy of items
- Understanding of the meaning of unit of measurement

Challenge

- Developing more advanced conceptual theories

38

The Rasch Model and Metrology – the Science of Measurement and its Application (towards RMT 2.0)

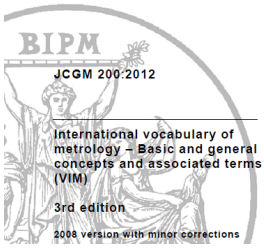
Metrological traceability

Uncertainty

2.41 (6.10)
metrological traceability

property of a **measurement result** whereby the result can be related to a reference through a documented unbroken chain of **calibrations**, each contributing to the **measurement uncertainty**

NOTE 1 For this definition, a 'reference' can be a definition of a **measurement unit** through its practical realization, or a **measurement procedure** including the measurement unit for a non-**ordinal quantity**, or a **measurement standard**.



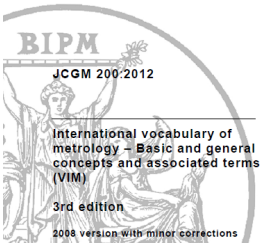
39

The Rasch Model and Metrology – the Science of Measurement and its Application (towards RMT 2.0)

Metrological traceability

Uncertainty

Precision based on actually available information
uncertainty expressed in a stable unit



2.26 (3.9)
measurement uncertainty
uncertainty of measurement
uncertainty

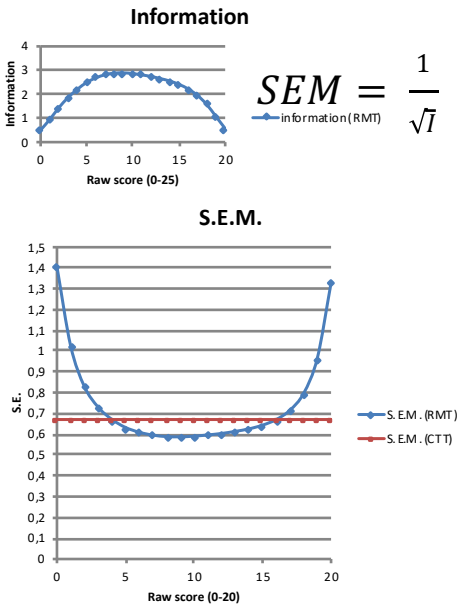
non-negative parameter characterizing the dispersion of the **quantity values** being attributed to a **measurand**, based on the information used

40

Uncertainty: Reliability and the S.E.M. in RMT and CTT

$$Rel = \frac{VAR_T}{VAR_X}$$

$$SEM = SDx \times \sqrt{1 - Rel}$$



$$SEM = \frac{1}{\sqrt{I}}$$

Example of actual data

	value (w/o extremes)
Cronbach's α	0.91 (0.88)
PSI	0.88 (0.87)
Person SD	2.24 (1.83)
Person VAR	5.00 (3.34)
n	2434

Expressing RMT's S.E.M. in the Reliability "Metric"

Idea:

Stating the S.E.M., which is expressed in an unfamiliar metric, at a given level in the more familiar reliability metric of 0 to 1

~ Local reliability
What would reliability be, if the sample distribution would have a particular observed variance and the SEM would be the same for all respondents.

Example of actual data

	value (w/o extremes)
Cronbach's α	0.91 (0.88)
PSI	0.88 (0.87)
Person SD	2.24 (1.83)
Person VAR	5.00 (3.34)
n	2434

42

Expressing RMT's S.E.M. in the Reliability "Metric"

$$Rel = 1 - \frac{SEM^2}{1}$$

Thissen (2000)

$$Rel = 1 - \frac{SEM^2}{VAR_x}$$

Assumption that observed variance = 1

Definition of the metric in IRT by setting the person true variance to 1

Locally SEM² might be bigger than 1 resulting in a nonsensical reliability < 0

Example of actual data

	value (w/o extremes)
Cronbach's α	0.91 (0.88)
PSI	0.88 (0.87)
Person SD	2.24 (1.83)
Person VAR	5.00 (3.34)
n	2434

43

Expressing RMT's S.E.M. in the Reliability "Metric"

$$Rel = 1 - \frac{SEM^2}{1 + SEM^2}$$

Ramsay (2000)

Assumption that true variance = 1
IRT: true variance = 1

$$Rel = 1 - \frac{SEM^2}{VAR_T + SEM^2}$$

RMT: item discrimination = 1
True variance is estimated

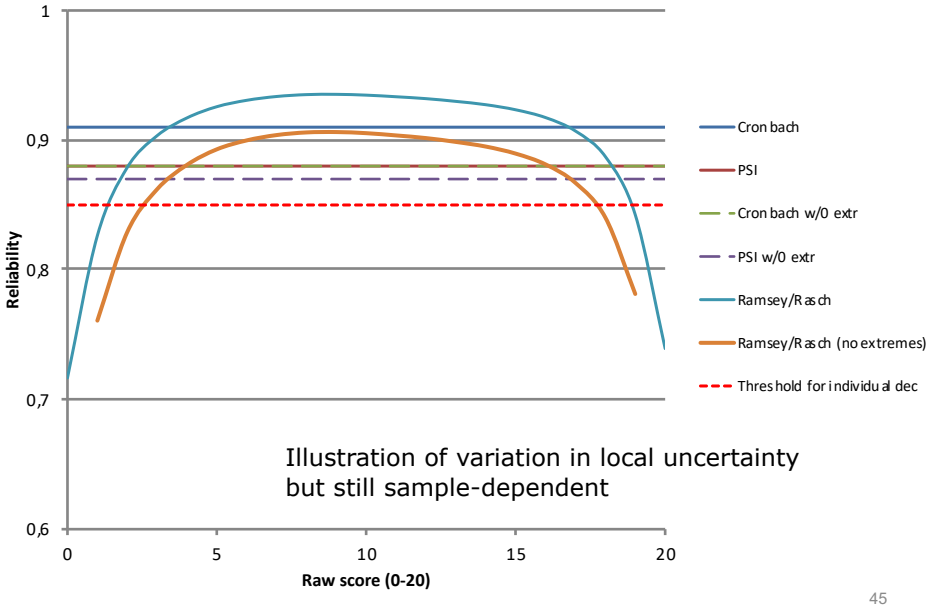
Ramsey/Rasch

Example of actual data

	value (w/o extremes)
Cronbach's α	0.91 (0.88)
PSI	0.88 (0.87)
Person SD	2.24 (1.83)
Person VAR	5.00 (3.34)
n	2434

44

Expressing RMT's S.E.M. in the Reliability "Metric"



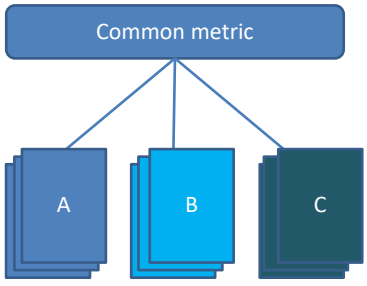
45

The Rasch Model and Metrology

Metrological traceability

Uncertainty

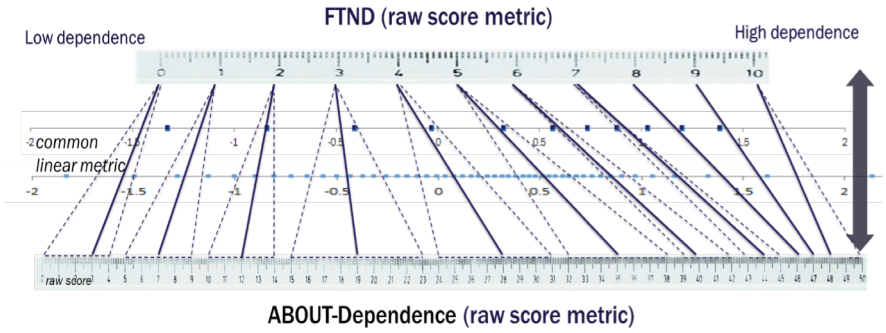
Precision based on actually available information
uncertainty expressed in a stable unit



Transcending instrumentation and application

46

Co-calibration of Dependence Instruments (Salzberger et al., 2019)



47

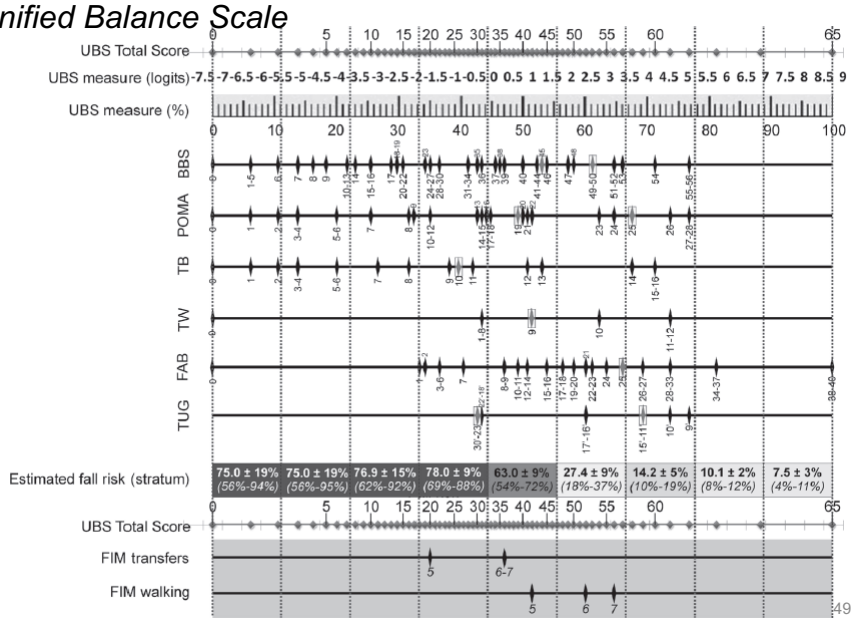
Enhanced Interpretation of Measurement

Raw score	Rasch composite measure		Most likely domain raw score			Domain EXTENT-OF-USE		Domain FEELING					Domain DOING					
	EXCL ECIG WML Rasch	POLY 0-100	DOING	FEELING	EXTENT	How soon after you woke up (01)	How long before going to bed (02)	strong desire (08)	difficult completely quit (06)	had to have one (10)	function normally (3)	hard to control the need/urge (13)	Use more than intended (09)	stop what you are doing (18)	Use in situations not supposed to (11)	sneak off to use product (16)	avoid an activity (17)
0	-2.585	0	0	0	0	>3 hrs	>3 hrs	NEVER	NEVER	NEVER	NEVER	NEVER	NEVER	NEVER	NEVER	NEVER	NEVER
1	-1.967	12	0	0	1	>3 hrs	>1-3 hrs	NEVER	NEVER	NEVER	NEVER	NEVER	NEVER	NEVER	NEVER	NEVER	NEVER
2	-1.587	19	0	1	1	>3 hrs	>1-3 hrs	RARELY	NEVER	NEVER	NEVER	NEVER	NEVER	NEVER	NEVER	NEVER	NEVER
3	-1.354	24	0	2	1	>3 hrs	>1-3 hrs	RARELY	NEVER	RARELY	NEVER	NEVER	NEVER	NEVER	NEVER	NEVER	NEVER
4	-1.185	27	0	3	1	>3 hrs	>1-3 hrs	RARELY	RARELY	RARELY	NEVER	NEVER	NEVER	NEVER	NEVER	NEVER	NEVER
5	-1.051	30	0	3	2	>1-3 hrs	>1-3 hrs	RARELY	RARELY	RARELY	NEVER	NEVER	NEVER	NEVER	NEVER	NEVER	NEVER
6	-0.938	32	0	4	2	>1-3 hrs	>1-3 hrs	SOMETIMES	RARELY	RARELY	NEVER	NEVER	NEVER	NEVER	NEVER	NEVER	NEVER
7	-0.839	34	0	5	2	>1-3 hrs	>1-3 hrs	SOMETIMES	RARELY	RARELY	RARELY	NEVER	NEVER	NEVER	NEVER	NEVER	NEVER
8	-0.750	35	0	6	2	>1-3 hrs	>1-3 hrs	SOMETIMES	RARELY	RARELY	RARELY	RARELY	NEVER	NEVER	NEVER	NEVER	NEVER
9	-0.669	37	0	6	3	>1-3 hrs	31-60 min	SOMETIMES	RARELY	RARELY	RARELY	RARELY	NEVER	NEVER	NEVER	NEVER	NEVER
10	-0.594	39	1	6	3	>1-3 hrs	31-60 min	SOMETIMES	RARELY	RARELY	RARELY	RARELY	RARELY	NEVER	NEVER	NEVER	NEVER
11	-0.525	40	1	7	3	>1-3 hrs	31-60 min	SOMETIMES	RARELY	SOMETIMES	RARELY	RARELY	RARELY	NEVER	NEVER	NEVER	NEVER
12	-0.460	41	2	7	3	>1-3 hrs	31-60 min	SOMETIMES	RARELY	SOMETIMES	RARELY	RARELY	RARELY	RARELY	NEVER	NEVER	NEVER
13	-0.399	42	2	8	3	>1-3 hrs	31-60 min	SOMETIMES	SOMETIMES	SOMETIMES	RARELY	RARELY	RARELY	RARELY	NEVER	NEVER	NEVER
14	-0.342	43	2	8	4	31-60 min	31-60 min	SOMETIMES	SOMETIMES	SOMETIMES	RARELY	RARELY	RARELY	RARELY	NEVER	NEVER	NEVER
15	-0.289	44	3	8	4	31-60 min	31-60 min	SOMETIMES	SOMETIMES	SOMETIMES	RARELY	RARELY	SOMETIMES	RARELY	NEVER	NEVER	NEVER
16	-0.238	45	3	9	4	31-60 min	31-60 min	SOMETIMES	SOMETIMES	SOMETIMES	SOMETIMES	RARELY	SOMETIMES	RARELY	NEVER	NEVER	NEVER
17	-0.189	46	3	9	5	16-30 min	31-60 min	SOMETIMES	SOMETIMES	SOMETIMES	SOMETIMES	RARELY	SOMETIMES	RARELY	NEVER	NEVER	NEVER
18	-0.142	47	4	9	5	16-30 min	31-60 min	SOMETIMES	SOMETIMES	SOMETIMES	SOMETIMES	RARELY	SOMETIMES	SOMETIMES	NEVER	NEVER	NEVER
19	-0.097	48	4	10	5	16-30 min	31-60 min	SOMETIMES	SOMETIMES	SOMETIMES	SOMETIMES	SOMETIMES	SOMETIMES	SOMETIMES	NEVER	NEVER	NEVER
20	-0.053	49	4	10	6	16-30 min	16-30 min	SOMETIMES	SOMETIMES	SOMETIMES	SOMETIMES	SOMETIMES	SOMETIMES	SOMETIMES	NEVER	NEVER	NEVER
21	-0.010	50	5	10	6	16-30 min	16-30 min	SOMETIMES	SOMETIMES	SOMETIMES	SOMETIMES	SOMETIMES	SOMETIMES	SOMETIMES	RARELY	NEVER	NEVER
22	0.032	51	5	11	6	16-30 min	16-30 min	SOMETIMES	MOST OF THE TIME	SOMETIMES	SOMETIMES	SOMETIMES	SOMETIMES	SOMETIMES	RARELY	NEVER	NEVER
23	0.072	51	5	11	7	6-15 min	16-30 min	SOMETIMES	MOST OF THE TIME	SOMETIMES	SOMETIMES	SOMETIMES	SOMETIMES	SOMETIMES	RARELY	NEVER	NEVER
24	0.113	52	5	11	8	6-15 min	6-15 min	SOMETIMES	MOST OF THE TIME	SOMETIMES	SOMETIMES	SOMETIMES	SOMETIMES	SOMETIMES	RARELY	NEVER	NEVER
25	0.152	53	5	11	9	0-5 min	6-15 min	SOMETIMES	MOST OF THE TIME	SOMETIMES	SOMETIMES	SOMETIMES	SOMETIMES	SOMETIMES	RARELY	NEVER	NEVER
26	0.191	54	5	11	10	0-5 min	0-5 min	SOMETIMES	MOST OF THE TIME	SOMETIMES	SOMETIMES	SOMETIMES	SOMETIMES	SOMETIMES	RARELY	NEVER	NEVER
27	0.229	54	6	11	10	0-5 min	0-5 min	SOMETIMES	MOST OF THE TIME	SOMETIMES	SOMETIMES	SOMETIMES	SOMETIMES	SOMETIMES	RARELY	RARELY	NEVER
28	0.266	55	6	12	10	0-5 min	0-5 min	MOST OF THE TIME	MOST OF THE TIME	SOMETIMES	SOMETIMES	SOMETIMES	SOMETIMES	SOMETIMES	RARELY	RARELY	NEVER
29	0.303	56	7	12	10	0-5 min	0-5 min	MOST OF THE TIME	MOST OF THE TIME	SOMETIMES	SOMETIMES	SOMETIMES	SOMETIMES	SOMETIMES	RARELY	SOMETIMES	NEVER
30	0.339	57	7	13	10	0-5 min	0-5 min	MOST OF THE TIME	MOST OF THE TIME	SOMETIMES	MOST OF THE TIME	SOMETIMES	SOMETIMES	SOMETIMES	RARELY	SOMETIMES	NEVER
31	0.374	57	8	13	10	0-5 min	0-5 min	MOST OF THE TIME	MOST OF THE TIME	SOMETIMES	MOST OF THE TIME	SOMETIMES	SOMETIMES	SOMETIMES	RARELY	SOMETIMES	RARELY
32	0.409	58	8	14	10	0-5 min	0-5 min	MOST OF THE TIME	ALL THE TIME	SOMETIMES	MOST OF THE TIME	SOMETIMES	SOMETIMES	SOMETIMES	RARELY	SOMETIMES	RARELY
33	0.442	59	9	14	10	0-5 min	0-5 min	MOST OF THE TIME	ALL THE TIME	SOMETIMES	MOST OF THE TIME	SOMETIMES	SOMETIMES	SOMETIMES	SOMETIMES	SOMETIMES	RARELY
34	0.476	59	9	15	10	0-5 min	0-5 min	MOST OF THE TIME	ALL THE TIME	MOST OF THE TIME	MOST OF THE TIME	SOMETIMES	SOMETIMES	SOMETIMES	SOMETIMES	SOMETIMES	RARELY
35	0.510	60	10	15	10	0-5 min	0-5 min	MOST OF THE TIME	ALL THE TIME	MOST OF THE TIME	MOST OF THE TIME	SOMETIMES	SOMETIMES	SOMETIMES	SOMETIMES	SOMETIMES	SOMETIMES
36	0.544	61	11	15	10	0-5 min	0-5 min	MOST OF THE TIME	ALL THE TIME	MOST OF THE TIME	MOST OF THE TIME	SOMETIMES	MOST OF THE TIME	SOMETIMES	SOMETIMES	SOMETIMES	SOMETIMES
37	0.577	61	11	16	10	0-5 min	0-5 min	MOST OF THE TIME	ALL THE TIME	MOST OF THE TIME	MOST OF THE TIME	MOST OF THE TIME	MOST OF THE TIME	SOMETIMES	SOMETIMES	SOMETIMES	SOMETIMES
38	0.610	62	12	16	10	0-5 min	0-5 min	MOST OF THE TIME	ALL THE TIME	MOST OF THE TIME	MOST OF THE TIME	MOST OF THE TIME	MOST OF THE TIME	MOST OF THE TIME	SOMETIMES	SOMETIMES	SOMETIMES
39	0.647	63	12	17	10	0-5 min	0-5 min	MOST OF THE TIME	ALL THE TIME	MOST OF THE TIME	ALL THE TIME	MOST OF THE TIME	MOST OF THE TIME	MOST OF THE TIME	SOMETIMES	SOMETIMES	SOMETIMES
40	0.683	63	13	17	10	0-5 min	0-5 min	MOST OF THE TIME	ALL THE TIME	MOST OF THE TIME	ALL THE TIME	MOST OF THE TIME	MOST OF THE TIME	MOST OF THE TIME	MOST OF THE TIME	SOMETIMES	SOMETIMES
41	0.721	64	14	17	10	0-5 min	0-5 min	MOST OF THE TIME	ALL THE TIME	MOST OF THE TIME	ALL THE TIME	MOST OF THE TIME	MOST OF THE TIME	MOST OF THE TIME	MOST OF THE TIME	MOST OF THE TIME	SOMETIMES
42	0.761	65	14	18	10	0-5 min	0-5 min	ALL THE TIME	ALL THE TIME	MOST OF THE TIME	ALL THE TIME	MOST OF THE TIME	MOST OF THE TIME	MOST OF THE TIME	MOST OF THE TIME	MOST OF THE TIME	SOMETIMES
43	0.808	66	15	18	10	0-5 min	0-5 min	ALL THE TIME	ALL THE TIME	MOST OF THE TIME	ALL THE TIME	MOST OF THE TIME	MOST OF THE TIME	MOST OF THE TIME	MOST OF THE TIME	MOST OF THE TIME	MOST OF THE TIME
44	0.859	67	16	18	10	0-5 min	0-5 min	ALL THE TIME	ALL THE TIME	MOST OF THE TIME	ALL THE TIME	MOST OF THE TIME	ALL THE TIME	MOST OF THE TIME	MOST OF THE TIME	MOST OF THE TIME	MOST OF THE TIME
45	0.919	68	16	19	10	0-5 min	0-5 min	ALL THE TIME	ALL THE TIME	ALL THE TIME	ALL THE TIME	MOST OF THE TIME	ALL THE TIME	MOST OF THE TIME	MOST OF THE TIME	MOST OF THE TIME	MOST OF THE TIME
46	0.993	69	17	19	10	0-5 min	0-5 min	ALL THE TIME	ALL THE TIME	ALL THE TIME	ALL THE TIME	MOST OF THE TIME	ALL THE TIME	ALL THE TIME	MOST OF THE TIME	MOST OF THE TIME	MOST OF THE TIME
47	1.091	71	17	20	10	0-5 min	0-5 min	ALL THE TIME	ALL THE TIME	ALL THE TIME	ALL THE TIME	ALL THE TIME	ALL THE TIME	ALL THE TIME	MOST OF THE TIME	MOST OF THE TIME	MOST OF THE TIME
48	1.234	74	18	20	10	0-5 min	0-5 min	ALL THE TIME	ALL THE TIME	ALL THE TIME	ALL THE TIME	ALL THE TIME	ALL THE TIME	ALL THE TIME	ALL THE TIME	MOST OF THE TIME	MOST OF THE TIME
49	1.489	79	19	20	10	0-5 min	0-5 min	ALL THE TIME	ALL THE TIME	ALL THE TIME	ALL THE TIME	ALL THE TIME	ALL THE TIME	ALL THE TIME	ALL THE TIME	ALL THE TIME	MOST OF THE TIME
50	1.943	86	20	20	10	0-5 min	0-5 min	ALL THE TIME	ALL THE TIME	ALL THE TIME	ALL THE TIME	ALL THE TIME	ALL THE TIME	ALL THE TIME	ALL THE TIME	ALL THE TIME	ALL THE TIME

48

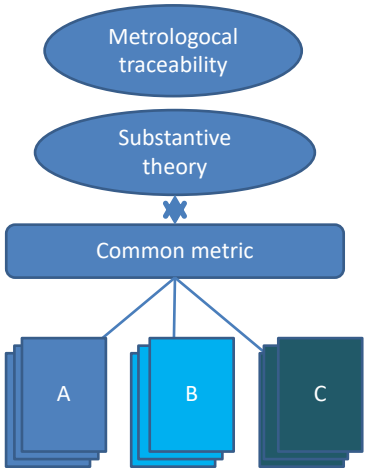
Common metric for Balance Scales (LaPorta et al., 2011)

Unified Balance Scale



49

Challenges for the Next 60 Years



Causal Rasch Model (Stenner et al., 2013)

Theory-based item hierarchy

Explicit, theory-based unit of measurement rather than emerging, implicit unit

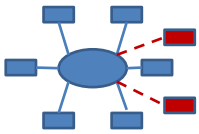
Transcending instrumentation and application

50

Advancing the Conceptual theory

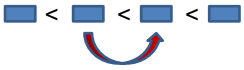
Qualitative theory of the attribute

- Suggests item content
- Updated based on empirical evidence
- *Descriptive theory*



Theory suggesting order of items (ordinal theory)

- General principle governing the order of items?
- Updated based on empirical evidence
- *(Partly) descriptive theory*

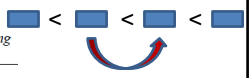


Quantitative theory of the attribute

- Specifies measurement mechanism that explains relationship between items (quantitative predictions with respect to item location)

51

Example: Measuring Cognitive Dissonance



weak

strong

levels

characteristics (cognitive, emotional)

Table 7.1 A theoretical construct map of cognitive dissonance in marketing

	Level of dissonance	Cognitive characteristics	Emotional characteristics
Weak cognitive dissonance	Very mild dissonance	Possibly not having had a sufficient look-around	Uneasiness, denial of being happy
	Mild dissonance	Lack of certainty about the purchase; possibly not enough information asked for	
	Noteworthy dissonance	Regretting the money spent on the product purchased	Denial of having a good feeling
	Moderate dissonance	Increased uncertainty whether purchase was sensible	Annoyance
Strong cognitive dissonance	Considerable dissonance	Expecting negative surprises during use of product; impression that product does not fit	Denial of pleasure given by the product in the future
	Strong dissonance	Wouldn't buy product again; feeling that they were spinning me a line	Unpleasant feeling when explaining purchase to friends; belief that product won't go down well with friends

52

Example: Measuring Cognitive Dissonance

Items

Table 7.2(cont.) *Twenty-eight items of CD in the original pool*

Level of dissonance	Item
Moderate dissonance	f14 **I wonder whether I should have spent less.** (Ich frage mich, ob ich nicht weniger ausgeben hätte sollen.) #
	f5 **I wonder whether I should have spent more to buy a better product.** (Ich frage mich, ob ich nicht für ein besseres Produkt mehr ausgeben hätte sollen.) #
	f4 **I wonder if I really need this product.** (Ich weiß gar nicht, ob ich das Produkt wirklich brauchen werde.) #
	f7 **I wonder whether I should have bought anything at all.** (Ich frage mich, ob ich überhaupt etwas kaufen hätte sollen.) #
	f9 **I am annoyed because once more I bought something I probably should not have bought.** (Ich ärgere mich, weil ich mir schon wieder etwas gekauft habe, was ich mir vielleicht gar nicht kaufen hätte sollen.) E
	f28 **I probably will have a look at a consumer magazine to be sure that the product I just bought performs better than alternative products.** (Ich werde wahrscheinlich eine Konsumenten-zeitschrift zur Hand nehmen, um sicher zu gehen, dass das eben gekaufte Produkt im Vergleich zu anderen Produkten gut liegt.)
	f27 **I will probably ask my family or friends whether they also think that I have made a good decision.** (Ich werde wahrschein-lich meine Familie oder Freunde fragen, ob sie auch meinen, dass ich mit diesem Kauf eine gute Entscheidung getroffen habe.)
Considerable dissonance	f15r **I am sure I will not experience negative surprises when using the product.*** (Ich bin mir sicher, dass ich beim Gebrauch des Produkts keine unangenehmen Überraschungen erleben werde.)
	f12[b] **Basically, the product does not fit me well.** (Im Grunde passt das Produkt gar nicht zu mir.)
	f26r **I am sure I will have a lot of fun with this product.*** (Ich bin sicher, dass ich mit den Leistungen dieses Produkts noch viel Freude haben werde.) E

Example: Measuring Cognitive Dissonance

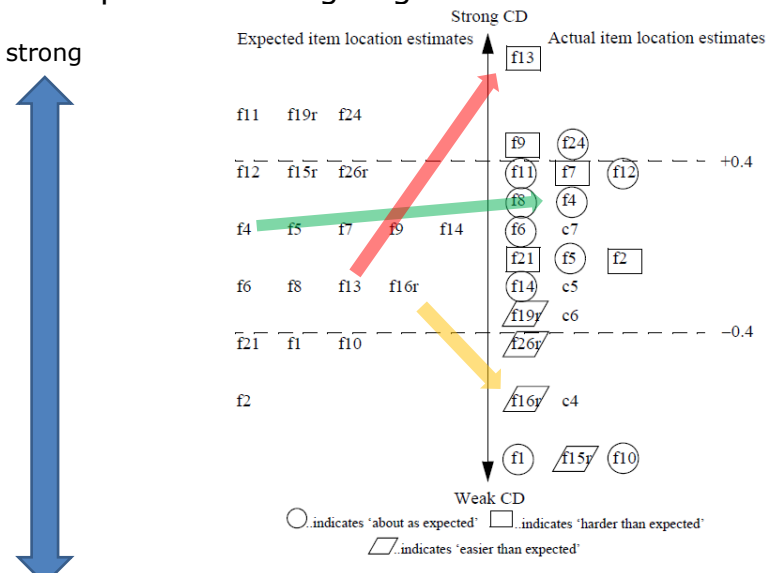


Figure 7.17 *Comparison of expected item endorsability and actual item location estimates (overall)*

54

Advancing Measurement in Health (the Social Sciences)

Improve definition of the concept to be measured and establish measurement standards (not instrument dependent)

"Thinking like physicists" (Stenner & Burdick, 2011)*

"Concept-driven measurement" (Cano et al., 2018)

Quantitative theory that exposes the measurement mechanism, causal Rasch models (Stenner et al., 2013)

Towards an explicit unit and measurement that transcends a particular instrument (Humphry, 2005; Humphry & Andrich, 2008; Humphry et al., 2014)

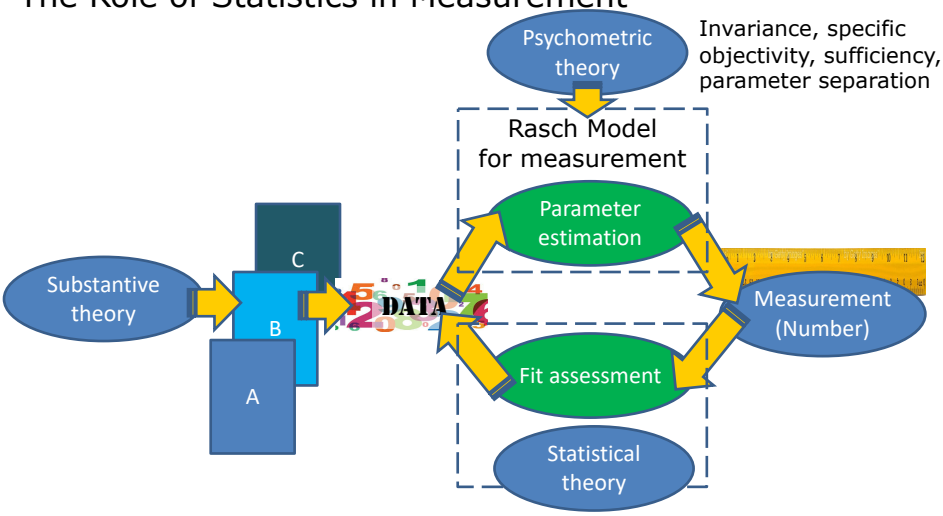
Linking metrology and psychometrics (Pendrill, 2017, 2014; Cano et al., 2016; Cano et al., 2017; Cano et al., 2018, 2019; Pendrill & Fisher, 2013)

-> Universal interpretation of measures (see Browne & Cano, 2019)

- *Reese, T.W. (1943). The Application of the Theory of Physical Measurement to the Measurement of Psychological Magnitudes, with Three Experimental Examples. Psychological Monographs,* ***55****, 1-89.*
  "The conclusion was drawn that none of the attempts at measurement, used so far by psychologists, meet the necessary criteria for fundamental measurement."

55

The Role of Statistics in Measurement



Fit assessment not to be based on – in the end arbitrary – cut-off values (e.g. p values), theory-informed decision making, but statistical evidence need to be accurate and trustworthy

56

Why RMT 2.0?

- Overcoming the threat of just "do[ing] so and so and so forth", of just "follow[ing] the form"
- Finding laws requires theorizing about laws

"Because of the success of science, there is a kind of a pseudo-science. Social science is an example of a science which is not a science. They follow the forms. You gather data, you do so and so and so forth, but they don't get any laws, they haven't found out anything. They haven't got anywhere – yet. Maybe someday they will, but it's not very well developed."

Richard Feynman, BBC Interview, 1981

57

58

References I

- Andersen, E.B. (1977). Sufficient Statistics and Latent Trait Models. *Psychometrika*, 42, 69-81.
- Andrich, D. (1978). A Rating Formulation for Ordered Response Categories. *Psychometrika*, 43 (4), 561-573.
- Andrich, D., & Kreiner, S. (2010). Quantifying response dependence between two dichotomous items using the Rasch model. *Applied Psychological Measurement*, *34*(3), 181-192.
- Andrich, D., & Hagquist, C. (2012). Real and artificial differential item functioning. Journal of Educational and Behavioral Statistics, 37(3), 387-416.
- Andrich, D., Humphry, S. M., & Marais, I. (2012). Quantifying local, response dependence between two polytomous items using the Rasch model. Applied Psychological Measurement, 36(4), 309-324.
- Andrich, D., Marais, I., & Humphry, S. (2012). Using a theorem by Andersen and the dichotomous Rasch model to assess the presence of random guessing in multiple choice items. Journal of Educational and Behavioral Statistics, 37(3), 417-442.
- Andrich, D. (2013). An expanded derivation of the threshold structure of the polytomous Rasch model that dispels any "threshold disorder controversy". Educational and Psychological Measurement, 73(1), 78-124.
- Andrich, D., & Marais, I. (2014). Person proficiency estimates in the dichotomous Rasch model when random guessing is removed from difficulty estimates of multiple choice items. Applied Psychological Measurement, 38(6), 432-449.
- Andrich, D., & Hagquist, C. (2015). Real and artificial differential item functioning in polytomous items. Educational and Psychological Measurement, 75(2), 185-207.
- Andrich, D. (2015). Components of variance of scales with a subscale structure using two calculations of coefficient alpha.(Pensamiento Educativo) Journal of Latin American Educational Research 52(2), 6-33.

References II

- Andrich, D. (2016). Controlling Response Dependence in the Measurement of Change using the Rasch model.
- Andrich, D. (2016). Components of Variance of Scales With a Bifactor Subscale Structure From Two Calculations of α. Educational Measurement: Issues and Practice, 35(4), 25-30.
- Andrich, D. (2017). Advances in social measurement: A Rasch measurement theory. In Perceived health and adaptation in chronic disease: Stakes and future challenge. In F. Guillemin, A. Leplège, S. Briançon, E. Spitz, and J. Coste (Eds.).Taylor and Francis.
- Browne, J. P., Cano, S. J. (2019). A Rasch Measurement Theory Approach to Improve the Interpretation of Patient-reported Outcomes. Medical care, 57, S18-S23.
- Cano, S. J., Vosk, T., Pendrill, L. R., & Stenner, A. J. (2016, November). On Trial: the Compatibility of Measurement in the Physical and Social Sciences. In Journal of Physics: Conference Series (Vol. 772, No. 1, p. 012025). IOP Publishing.
- Cano, S., Pendrill, L.R., Barbic, S., Fisher, W.P. (2017). Patient-centred outcome metrology for healthcare decision-making. In Joint IMEKO TC1-TC7-TC13 Symposium:"Measurement Science Challenges in Natural and Social Sciences" Rio de Janeiro, Brazil, from July 31st to August 3rd 2017.
- Cano, S., Pendrill, L.R., Melin, J., Köbe, T., Fisher, W.P. (2018). Metrology for the Social Sciences: A case for Rasch Measurement Theory (not Rasch Analysis), IOMW 2018 New York.
- Cano, S. J., Pendrill, L. R., Melin, J., Fisher Jr, W. P. (2019). Towards consensus measurement standards for patient-centered outcomes. Measurement, 141, 62-69.

References III

- Christensen, K.B., Makransky, G., Horton, M.: Critical Values for Yen's Q3: Identification of Local Dependence in the Rasch Model Using Residual Correlations. Applied Psychological Measurement 2017, 41:178-194.
- Draxler, C. (2010). Sample size determination for Rasch model tests. Psychometrika, 75(4), 708-724.
- Draxler, C., & Alexandrowicz, R. W. (2015). Sample size determination within the scope of conditional maximum likelihood estimation with special focus on testing the Rasch model. psychometrika, 80(4), 897-919.
- Ewing, M. T., Salzberger, T., & Sinkovics, R. R. (2005). An alternate approach to assessing cross-cultural measurement equivalence in advertising research. Journal of Advertising, 34(1), 17-36.
- Hagell, P., & Westergren, A. (2016). Sample Size and Statistical Conclusions from Tests of Fit to the Rasch Model According to the Rasch Unidimensional Measurement Model (Rumm) Program in Health Outcome Measurement. Journal of applied measurement, 17(4), 416-431.
- Hagquist, C., & Andrich, D. (2015). Determinants of artificial DIF–a study based on simulated polytomous data. Psychological Test and Assessment Modeling, 57(3), 342-376.
- Hobart, J. C., Cano, S. J., Warner, T. T., & Thompson, A. J. (2012). What sample sizes for reliability and validity studies in neurology?. Journal of neurology, 259(12), 2681-2694.
- Humphry, S. (2005). Maintaining a common arbitrary unit in social measurement (Doctoral dissertation, Murdoch University).

References IV

- Humphry, S. M., Andrich, D. (2008). Understanding the unit in the Rasch model. Journal of applied measurement, 9(3), 249-264.
- Humphry, S., Heldsinger, S., Andrich, D. (2014). Requiring a consistent unit of scale between the responses of students and judges in standard setting. Applied Measurement in Education, 27(1), 1-18.
- Kubinger, K. D., Rasch, D., Yanagida, T. (2009). On designing data-sampling for Rasch model calibrating an achievement test. Psychology Science Quarterly, 51(4), 370-84.
- Kubinger, K. D., Rasch, D., & Yanagida, T. (2011). A new approach for testing the Rasch model. Educational Research and Evaluation, 17(5), 321-333.
- Marais, I., & Andrich, D. (2008). Formalizing dimension and response violations of local independence in the unidimensional Rasch model. *J Appl Meas*, *9*(3), 200-15.
- Marais, I. (2013). Local dependence. In Rasch models in health Edited by Christensen KB, Kreiner S, Mesbah M. London, UK: Wiley-ISTE Ltd. : 111-130.
- Maul, A. (2017). Rethinking Traditional Methods of Survey Validation. Measurement: Interdisciplinary Research and Perspectives, 15(2), 51-69.
- Michell, J. (1999). Measurement in Psychology – a Critical History of a Methodological Concept, Cambridge: Cambridge University Press.
- O'Connell, J. (1993). Metrology: The creation of universality by the circulation of particulars. Social studies of science, 23(1), 129-173.
- Olsbjerg, M., & Christensen, K. B. (2015). Modeling local dependence in longitudinal IRT models. *Behavior research methods*, *47*(4), 1413-1424.
- Pendrill, L. R., & Fisher Jr, W. P. (2013). Quantifying Human Response: Linking metrological and psychometric characterisations of Man as a Measurement Instrument. In Journal of Physics: Conference Series (Vol. 459, No. 1, p. 012057). IOP Publishing.

References V

- Pendrill, L. (2014). Man as a measurement instrument. NCSLI Measure, 9(4), 24-35.
- Pendrill, L. R. (2017). Assuring measurement quality in person-centred healthcare. Measurement Science and Technology.
- Rasch, G. (1960). Probabilistic Models for Some Intelligence and Attainment Tests, Copenhagen: Danish Institute for Educational Research, expanded edition (1980) with foreword and afterword by B.D. Wright. Chicago: The University of Chicago Press.
- Rasch, G. (1961). On General Laws and the Meaning of Measurement in Psychology. Berkeley Symposium on Mathematical Statistics and Theory of Probability, Vol. IV, 20/6–30/7 1960, Berkeley: University of California Press, pp. 321-333.
- Rasch, G. (1977). On Specific Objectivity: an Attempt at Formalizing the Request for Generality and Validity of Scientific Statements. *Danish Yearbook of Philosophy*, 14, 58-93.
- Reese, T.W. (1943). The Application of the Theory of Physical Measurement to the Measurement of Psychological Magnitudes, with Three Experimental Examples. *Psychological Monographs*, 55, 1-89.
- Stenner, A. J., & Burdick, D. S. (2011). Can psychometricians learn to think like physicists?. Measurement, 9(1), 62-63.
- Stenner, A. J., Fisher Jr, W. P., Stone, M. H., & Burdick, D. S. (2013). Causal Rasch models. Frontiers in psychology, 4, 1-14.